

THIS WEEK

EDITORIALS

WORLD VIEW The genome leak is approaching — are you ready? **p.139**

STARFISH Hot arms jettisoned to survive heating **p.140**



ASTRONOMY Stellar dust trap could explain planet formation **p.141**

No dishonour in depression

The stigma associated with mental illness discourages investment in finding cures — even though the burden of the disorders on society is immense.

Comedian and writer Ruby Wax, a regular on British television, has clinical depression. In her book published last week, *Sane New World* (Hodder & Stoughton, 2013), she describes her struggles with different therapies and her fear of being ‘found out’. She is not alone. A 2010 survey in Europe revealed that 38% of people had a diagnosed mental disorder — including 7% with major depression. The proportion is likely to be similar in all populations, even in Africa, where psychiatric disease barely features on the health agenda.

The stigma attached to such disorders means that many people do not admit to their illness. The same stigma discourages investment, so that research funding is not proportional to the distress these disorders cause. Why lobby for better treatments for depression or schizophrenia when there are ‘real’ diseases out there, such as cancer?

Wax has been through the catalogue of available therapies and says that she has settled on an approach known as ‘mindfulness’, which helps to keep her depression under control. It may seem that the various therapies are inadequate, given that initial treatment of depression fails in 60% or more of cases. It is true that more treatment options are badly needed. Yet evidence-based cognitive behavioural therapies and drugs already developed by the pharmaceutical industry can work splendidly for long periods — if they are given to the right patients.

How do you recognize the right patients? Treatment decisions tend to be based on the preferences of physicians or their patients, often with a missionary zeal that gives no credence to the idea that a personalized approach would be more appropriate. Some hold that drugs have unacceptable side effects, whereas others say that cognitive therapy wastes time if the depressed brain is not first chemically lifted. It is becoming increasingly common to offer patients both treatments at once in the belief that drugs can prepare the brain to respond to cognitive therapy. That may be so, but it is also possible that the improved response rates are simply the result of catching two different populations.

The situation would improve drastically if simple tests could be developed to predict treatment outcome. Many exploratory clinical trials are now under way to search for biomarkers in genes or in the brain itself that might be predictive. This week sees the description of the first potential biomarker for discriminating between responders and non-responders to drugs or cognitive therapy in major depressive disorder (C. L. McGrath *et al.* *JAMA Psychiatry* <http://dx.doi.org/10.1001/jamapsychiatry.2013.143>; 2013).

The study, led by neurologist Helen Mayberg of Emory University in Atlanta, Georgia, used positron emission tomography (PET) scans to measure metabolic activity in various brain regions of people with untreated depression (see also *Nature* <http://doi.org/mtc>; 2013). Patients were randomized into groups and treated for 12 weeks with either a commonly used antidepressant drug or cognitive behaviour therapy. The study’s results were clear-cut. Below-average activity in a brain area called the right anterior insula — which is linked with depression-relevant behaviours such as emotional self-awareness and decision-making

— was associated with the patient showing a good response to cognitive behavioural therapy and a poor response to the drug. Above-average insula activity was predictive of the opposite.

This potential biomarker must still be tested in prospective clinical trials, which will assign patients to a treatment on the basis of their insula activity. It may fail. But if the biomarker comes up trumps, it could be transformative for many patients who would not have to endure two or three months of treatment trial and error.

“The stigma attached to mental disorders will fade when treatment becomes more effective.”

If attitudes to mental illness do not change, even a successful biomarker of this type will have a hard time being accepted by health systems that foot the bills. Unlike a simple blood test, a PET scan is inconvenient because not all physicians have easy access to the technology and, at up to US\$2,000 a shot, the procedure is not cheap. Although expensive treatments for other diseases and arguments about how to fund them are nothing new, this rational debate is harder for mental illnesses because of the irrational stigma that is attached to them.

Fifty years ago, the stigma surrounding cancer meant that physicians would sometimes lie to patients about the diagnosis from kindness. That has now faded because cancer is not always the death sentence it once was — thanks in part to the development of biomarkers that guide therapy. The stigma attached to mental disorders will also fade when treatment becomes more effective. But to break out of a vicious circle of underinvestment in a stigmatized disease area will require continued effort to get the problem recognized. This is a good week for that. ■

Track the trackers

Oversight and public debate about access to personal data are crucial to preserving privacy.

Access to phone records is all over the news at the moment, following media revelations of massive snooping by US intelligence agencies. Collecting phone records is in itself nothing new, and is legitimate — scientists have long done so for research. What is unprecedented is the government weaving together multiple huge data sets for secret state surveillance.

When researchers obtain electronic records of millions of people’s calls for social science — including data on who called whom, when and from where, but not actual conversations — they must first make a strong case for why they need that information, and must comply with multiple layers of oversight and safeguards such as anonymization of

numbers. The same goes for research use of health records or any other private personal data. And there is a good reason: to protect privacy.

Privacy concerns are at the heart of the uproar over how the US National Security Agency (NSA) has secretly required telephone companies to hand over similar phone records on almost every US resident. The US government is also vacuuming up billions of e-mails and other Internet communications from traffic outside the United States — all in the name of law enforcement and the war on terror.

What is perhaps most concerning, apart from the mind-boggling scale of the snooping, is that until last week, the very existence of these programmes was secret. Since the revelations, US President Barack Obama has defended this secrecy, on the grounds that if terrorists knew that the government was monitoring phones and the Internet, they would seek ways around the surveillance. But most terrorists probably take that as a given and — unlike most ordinary citizens — already use encryption and other techniques to secure and obfuscate communication. It is a poor excuse for a lack of transparency and public oversight of such snooping. Obama asked Americans to trust the government, but history shows that ‘trust us’ is not good enough.

The revelations seem to vindicate many of the conclusions and recommendations of a 2008 report by the US National Research Council (NRC) — *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment* (go.nature.com/bs00ux). That report addressed privacy issues raised by the Total Information Awareness programme, a research effort launched by the US Defense Advanced Research Projects Agency in 2002 to develop data mining and other technologies to link and search disparate databases, for example to try to identify suspicious patterns to detect and track terrorists.

After much controversy, that programme had its funding removed by Congress in 2003. But as the NRC report noted, this was probably a pyrrhic victory for civil liberties. It removed a focused programme subject to congressional oversight and public debate that would determine appropriate uses and safeguards. Instead, much the same work has continued in agencies across government, including the NSA, with less oversight. The report warned that this was “likely to result in little security and, ultimately, brittle privacy protection”. How right it was.

Privacy matters. Yet last week, many defenders of snooping on private individuals sought to play down its significance. Several, including UK foreign secretary William Hague, trotted out tired fallacies, including that people who have nothing to hide have nothing to fear. That has long been debunked by academics; the idea is based on a misconception of what privacy is about.

Privacy is a human right, and is essential if people are to develop autonomy. It is central to freedom of expression and association, and to preventing abuse of personal information. There are numerous examples of misuse of private data by agencies and law enforcement, including intimidation, selective character assassination, repression of dissent and wrongful arrest. Privacy is a cornerstone of a free and creative society, and is an essential defence against unwarranted social control.

Government officials in the United States and elsewhere should find the NRC report and read it carefully. It calls for “robust, independent oversight” of government data mining and surveillance to “mine the miners and track the trackers”. Some data could help security efforts, the report says, but it notes that many security experts have misgivings. They question the feasibility and reliability of data mining to look for and track terrorists in massive data sets, and they raise concerns about the risk of law-abiding individuals and companies being falsely targeted.

Such surveillance is not unique to the United States. In April, a report by the United Nations’ Human Rights Council warned that many countries worldwide, including democracies, are increasingly allowing intelligence and law-enforcement agencies to deploy indiscriminate and extensive surveillance of communications. That weakens or removes safeguards such as justification of individual cases of surveillance, and oversight by a neutral judicial body.

As the World View on page 139 shows, privacy and what it means in the digital age is an increasingly crucial question in the era of big data. A grown-up and open debate is needed, with trust on all sides. It has not started well. ■

“Obama asked Americans to trust the government, but history shows that ‘trust us’ is not good enough.”

Young upstarts

Lucrative prizes emulating the Nobels bring welcome money and publicity for science.

When a theoretical physicist who has worked on quantum field and string theory calls attention to an “interesting experiment”, the experiment deserves notice. This is particularly true when that experiment is an attempt to deliver a little Hollywood glamour to physics, with an Oscars-style ceremony and gigantic cash prizes.

The US\$3-million Fundamental Physics Prize is indeed an interesting experiment, as Alexander Polyakov said when he accepted this year’s award in March. And it is far from the only one of its type. As a News Feature on page 152 discusses, a string of lucrative awards for researchers have joined the Nobel Prizes in recent years. Many, like the Fundamental Physics Prize, are funded from the telephone-number-sized bank accounts of Internet entrepreneurs. These benefactors have succeeded in their chosen fields, they say, and they want to use their wealth to draw attention to those who have succeeded in science.

What’s not to like? Quite a lot, according to a handful of scientists quoted in the News Feature. You cannot buy class, as the old saying goes, and these upstart entrepreneurs cannot buy their prizes the prestige of the Nobels. The new awards are an exercise in self-promotion for those behind them, say scientists. They could distort the

meritocracy of peer-review-led research. They could cement the status quo of peer-reviewed research. They do not fund peer-reviewed research. They perpetuate the myth of the lone genius.

The goals of the prize-givers seem as scattered as the criticism. Some want to shock, others to draw people into science, or to better reward those who have made their careers in research. Several want to show that leading scientists can attain the lifestyles of financiers and footballers.

As *Nature* has pointed out before, there are some legitimate concerns about how science prizes — both new and old — are distributed. The Breakthrough Prize in Life Sciences, launched this year, takes an unrepresentative view of what the life sciences include (see *Nature* 494, 402; 2013). But the Nobel Foundation’s limit of three recipients per prize, each of whom must still be living, has long been outgrown by the collaborative nature of modern research — as will be demonstrated by the inevitable row over who is ignored when it comes to acknowledging the discovery of the Higgs boson. The Nobels were, of course, themselves set up by a very rich individual who had decided what he wanted to do with his own money. Time, rather than intention, has given them legitimacy.

As much as some scientists may grumble about the new awards, the financial doping that they bring to research and the wisdom of the goals behind them, two things seem clear. First, most researchers would accept such a prize if they were offered one. Second, it is surely a good thing that the money and attention come to science rather than go elsewhere. It is fair to criticize and question the mechanism — that is the culture of research, after all — but it is the prize-givers’ money to do with as they please. It is wise to accept such gifts with gratitude and grace. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunjv



Be prepared for the big genome leak

It is only a matter of time until idealism sees the release of confidential genetic data on study participants, says Steven E. Brenner.

Most people in the United States could soon know someone whose genome is held in a research database. Concerns are growing about our ability to properly control access to that information. Also growing among some scientists is the feeling that restricting access to genomic data fetters research. How long will it be until an idealistic and technically literate researcher deliberately releases genome and trait information publicly in the name of open science?

Both the open-access literature and the open-source software movements began with idealists. It seems inevitable that there will be a major leak of genome information in the near future. Individual scientists, institutions and funders should consider now how they will react when this happens.

Some studies already gather the genetic data of more than 50,000 individuals in a single analysis. Although this information is supposed to be highly protected, it is disseminated to various institutions that have inconsistent security and privacy standards. In practice, data protection often comes down to individual scientists. Once leaked, these data would be virtually impossible to contain.

What harm would come from a leak of personal and genomic data? The consent form for the Personal Genome Project (PGP) — which makes no attempt to keep genetic information secret — offers a guide. It lists a range of adverse consequences, from revealing non-paternity to being framed with synthesized DNA planted at a crime scene.

Most research genome data are de-identified, but given progress in re-identification and commercial genetic databases, will they stay that way? De-anonymized genomic data would be most likely to reveal health conditions relevant to the study for which they were collected. The effects might be uncomfortable but would probably reveal less than a typical Google search history. So far, no PGP participant who released genomes and traits has experienced adverse consequences that have been reported to the Institutional Review Board. In the longer term, the risk of harm may rise as our understanding of genetic variation increases.

Then there is the public outcry a genome breach might incite. The public often has an exaggerated perception of the links between genes and personal traits. Lacking contextual information, research participants could wonder whether their own genomes had been leaked and dread implausibly dire consequences.

Thus a genome leak might lead to a backlash. Volunteers might withdraw from research studies and refuse to join new ones. Research might even be subject to moratoriums and prohibitive restrictions. The harm to genetic research could be great, and study participants could be unsettled.

What can be done? Two extreme options offer appealing simplicity. One is for research projects to incorporate unrestricted data release from the outset. This option should be offered more broadly owing to the certainty and research benefits it offers. However, would enough people be willing to share so openly? The second option would be to lock down genomes so tightly that they are virtually impossible to steal, for example by only allowing analyses on central computers through restricted interfaces. Although useful as an alternative, this system would stymie research were it to become the exclusive means of access to data, but it would still remain vulnerable to ingenious ways of eliciting inappropriate genetic information.

Neither option is comprehensively workable, which means that the question is not how to prevent a leak but how to mitigate the fall-out. This requires some specific steps, as well as progress in adapting

concepts already used elsewhere in biological research and in applying principles proposed by groups such as the Presidential Commission for the Study of Bioethical Issues in Washington DC.

Funders should develop rapid mechanisms for notifying study participants, governments and the media when breaches occur and provide informed guidance about scope and probable consequences for those affected. This would require recontacting research participants to warn those whose data were leaked and, implicitly, to calm others whose data remain secure. More research is needed about the possible harm of such leaks to better inform and protect research participants before and after leaks occur.

We should also take steps to minimize the frequency and extent of future genome leaks. Institutions could establish uniform protocols and reviews to ensure the safety of protected genomic data. All researchers using restricted genomic data should be trained regarding the ethics of and the technologies involved in protecting human data. Technical and legal strategies should be proactively deployed to help limit dissemination of leaked data to those who furtively hunt for them.

Augmented legal protections could reduce the harm from inappropriate use of such data. In the meantime, we need to address a quandary: research with leaked data would undoubtedly speed immediate scientific progress, but should scientists exploit them?

Most importantly, we must ensure that the necessary discussion about the risks of a genome leak is balanced with information about the tremendous benefits that collected genetic information has for all of us. Although the acceleration and promise of genomics makes a leak inevitable, it also guarantees medical progress. ■ [SEE EDITORIAL P.137](#)

Steven E. Brenner is a Professor at the University of California, Berkeley.
e-mail: brenner@compbio.berkeley.edu

THE QUESTION IS NOT
HOW TO PREVENT
A LEAK
BUT HOW TO
MITIGATE THE
FALL-OUT.

➔ [NATURE.COM](#)
Discuss this article
online at:
go.nature.com/oybzqm

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

DEVELOPMENT

Birds' mysterious missing penises

The development of chicken penises is cut short by signals that prompt cell death, a finding that could help to explain why 97% of bird species have little or no phallus despite reproducing by internal fertilization.

Researchers led by Martin Cohn at the University of Florida in Gainesville cut tiny windows into eggs to compare developing chickens, which lack phalluses, with ducks, whose penises can be half as long as their bodies.

Chicken embryos began to form penises, but these shrank midway through development. The researchers pinned the cause on elevated levels of a protein called Bmp4, which promotes cell death, at the tip of the organ.

The loss of penises may have been a by-product of the evolution of other features, such as beak shape, which are also influenced by Bmp proteins, the authors suggest.

Curr. Biol. <http://dx.doi.org/10.1016/j.cub.2013.04.062> (2013)

For a longer story on this research, see go.nature.com/1mgn9w

ANIMAL BEHAVIOUR

Sea stars shed too-hot arms

Sea stars may use their arms to keep their central cores cool when high



temperatures threaten their survival.

Sylvain Pincebourde at the University of Tours, France, and his colleagues kept ochre sea stars (*Pisaster ochraceus*, pictured) under conditions that mimicked the sweltering temperatures to which the organisms can be exposed at low tides. Most sea stars died if their core temperatures exceeded 35°C. Individuals that survived the heating generally had arms that were hotter than their cores,

possibly because the creatures reroute body fluids into their central cavities to cool down. When their cores warmed, sea stars shed arms — consistently losing the hottest one first. *J. Exp. Biol.* 216, 2183–2191 (2013)

CANCER BIOLOGY

Immunity let loose

Experimental therapies that unleash the immune system to fight cancer, by blocking

'checkpoint inhibitors', continue to show promise in early clinical trials.

Immune checkpoint inhibitors prevent autoimmunity, and can rein in the immune response against tumours. Antoni Ribas at the University of California, Los Angeles, and his colleagues tested lambrolizumab, a compound that blocks a checkpoint inhibitor called PD-1, in 135 people with advanced melanoma. Tumours shrank by at least 30% in 38% of the patients.



CLIMATE SCIENCE

Reindeer keep the ground cool

Reindeer herding practices and their effect on vegetation in northern Scandinavia may influence when snow melts in spring.

Tall, dense shrubs can hasten snow melt in the tundra. As more branches protrude over packed snow, less sunlight is reflected off the bright surface and more heat is absorbed by the ground.

A team led by Juval Cohen at the Finnish Meteorological Institute in Helsinki used satellite observations to examine the cover of vegetation

and snow in northern Scandinavia. In Finland, where reindeer typically graze on the tundra throughout the year, snow melt begins later. In inland Norway, where the pastures are left ungrazed during summer and vegetation is taller and more abundant, the snow melts earlier.

More intense reindeer grazing could delay snow melt and reduce ground heating during spring in the rapidly warming tundra, the authors suggest.

Remote Sens. Environ. 135, 107–117 (2013)

QUIRIN SCHIERMER

ED RESCHKE/PETER ARNOLD WORLDWIDE/GETTY IMAGES

In a separate study, Jedd Wolchok at Memorial Sloan-Kettering Cancer Center in New York and his colleagues, treated 86 people with advanced melanoma using two compounds: nivolumab, also a PD-1 inhibitor, and ipilimumab, an approved drug that blocks a checkpoint inhibitor called CTLA-4. Tumours shrank by at least half in 40% of the patients.

N. Engl. J. Med. <http://dx.doi.org/10.1056/NEJMoa1305133>; <http://dx.doi.org/10.1056/NEJMoa1302369> (2013)

ASTRONOMY

Trap holds protoplanet dust

Dust particles spotted around a young star support an idea about how planets are born.

Planet formation is a paradox: according to standard theory, dust grains orbiting newborn stars should spiral into those stars rather than accrete to form planets. Astronomers have suggested that there are regions, or 'pressure bumps', where density and pressure gradients trap particles long enough to allow them to clump together.

A team led by Nienke van der Marel at Leiden University in the Netherlands has observed such a trap around the star Oph IRS 48 located about 120 parsecs from Earth. The Atacama Large Millimeter/submillimeter Array in Chile detected a crescent-shaped cluster on one side of the star — probably a reservoir of coalescing dust grains (pictured as an artist's impression).

Science 340, 1199–1202 (2013)

ENZYME ENGINEERING

Serendipity outstrips design

Accelerated evolution of an artificial enzyme improved its activity several-thousand fold, owing to unexpectedly extreme remodelling of its active site.

A team led by Donald Hilvert and Nenad Ban at the Swiss Federal Institute of Technology in Zurich optimized a computationally designed enzyme with several rounds of random mutagenesis and screening. The activity levels of the evolving enzyme eventually approached those of natural enzymes, but the protein no longer catalysed its reaction using the machinery the researchers had intended. An amino-acid residue installed to help rearrange bonds was abandoned for one that emerged at another location in the active site. Such swaps could be important in natural-enzyme evolution and design efforts, the authors say. *Nature Chem. Biol.* <http://dx.doi.org/10.1038/nchembio.1276> (2013)

MOLECULAR BIOLOGY

Boosting plant defence

The discovery of a gene that regulates the effects of the plant hormone jasmonic acid might lead to ways to increase pest resistance in crops, without hindering their growth.

Jasmonic acid helps plants to fend off insects and pathogens; it also regulates aspects of plant development, including fertility and fruit ripening. Daoxin Xie of Tsinghua University in Beijing and his colleagues identified a gene called *JAV1* in the model plant *Arabidopsis thaliana* that suppresses several responses triggered by jasmonic acid. The *JAV1* protein was degraded when insects or fungi attacked. Silencing *JAV1* boosted plant resistance to disease, but had no adverse effect on fertility or other developmental processes. *Mol. Cell* 50, 506–517 (2013)

COMMUNITY CHOICE

The most viewed papers in science

CHEMISTRY

One polymer with multiple forms



An unusual material can switch between polymers from two different classes with the addition of light.

Da-Hui Qu, He Tian and their colleagues at the East China University of Science and Technology in Shanghai combined two types of molecules. Cyclodextrins form non-covalent complexes to yield supramolecular polymers, whereas coumarins form covalent bonds with each other under one wavelength of light, and release those bonds under another. The material that the researchers created could go from a supramolecular polymer to a covalent polymer and back again with the addition of light; adding a detergent produced a reversible hydrogel. Substances with switchable properties can combine advantages of distinct polymers in a single platform, the authors say.

Langmuir 29, 5345–5350 (2013)

PALAEONTOLOGY

Big lizard among mammals

A giant, plant-eating lizard successfully competed with mammals about 40 million to 36 million years ago.

Researchers led by Jason Head at the University of Nebraska-Lincoln identified the lizard in a diverse assemblage of fossils collected in Myanmar. The teeth and jaws of the creature revealed that it was a plant-eater, and at an estimated 27 kilograms, it was one of the largest animals in the area. The researchers dubbed the species — which was almost twice the length of any living herbivorous lizard — *Barbaturex morrisoni* after the singer Jim Morrison, who famously proclaimed himself the lizard king.

Reptiles need external heat to keep their bodies warm, so the hotter temperatures of past climates could have allowed the large lizards to survive, the authors say. *Proc. R. Soc. B* 280, 20130665 (2013)

PLANT SCIENCES

Tomatoes make tubers

Boosting levels of a hormone in tomato plants (*Solanum lycopersicum*) causes them to make tubers, like their sibling species the potato (*Solanum tuberosum*).

Yuval Eshed at the Weizmann Institute of Science in Rehovot, Israel, Eliezer Lifschitz at the Israel Institute of Technology in Haifa and their team engineered tomatoes to have high levels of a cytokinin, a type of hormone found in all plants. The tomatoes formed tiny tubers (pictured) at the base of leaves along their stems, where cells divide and levels of hormones fluctuate.

Giving potato plants the hormone in culture also elicited small spuds along plant stems. The authors suggest that a simple, common mechanism might prompt tubers in other species.

Curr. Biol. <http://dx.doi.org/10.1016/j.cub.2013.04.061> (2013)

► NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch



TAMAR EVIATAR-REBAK

ESO/L. CALÇADA



SEVEN DAYS

The news in brief

POLICY

Gun research

The US Institute of Medicine (IOM) has recommended a broad firearms-research agenda for the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. The 5 June IOM report poses questions on topics from the value of background checks to the influence of violent media and games on gun-related violence. The CDC requested the report after US President Barack Obama made an order in January for it to resume research on gun violence, which the agency had stopped in 1996 when Congress forbade it from using money to “advocate or promote” gun control. See go.nature.com/rufroe for more.

Brazil emissions

Brazil’s greenhouse-gas emissions fell by nearly 39% between 2005 and 2010, according to an inventory released on 5 June by the country’s government. The sharp drop was entirely due to falling rates of deforestation. However, it was tempered by rising emissions from the agriculture and energy sectors — a concern if Brazil wants to cut emissions further. Overall, the country is on track to meet goals announced at the 2009 United Nations climate summit in Copenhagen. See go.nature.com/xll3ht for more.

Watchdog backs off

The US federal office that punishes breaches in human-research protections said on 5 June that it would not issue sanctions over a controversial university study on how best to treat premature infants with oxygen. In March, the Office for Human Research Protections had said that investigators in the study,

which was overseen by the University of Alabama at Birmingham, did not sufficiently inform parents of the risks to infants in the trial. In its letter last week, the office said that the issues raised were complex enough to require new guidance for future studies.

Grey-wolf revival

Grey wolves (*Canis lupus*), once on the verge of extinction, have recovered sufficiently to be removed from the federal list of threatened and endangered species, the US Fish and Wildlife Service (FWS) proposed on 7 June. That would return management of the animals to state wildlife agencies. The FWS says that wolves are expanding their range and have exceeded

population targets by up to 300%, but environmentalists fear that state hunting initiatives will inhibit that recovery. The FWS proposal would maintain federal protection for the Mexican wolf in the southwest.

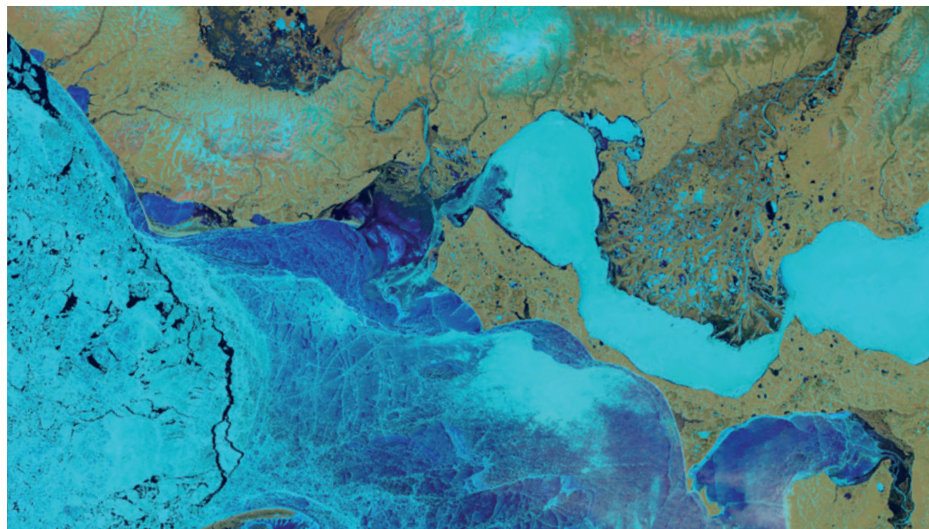
Montreal accord

China agreed on 8 June that it will “work together” with the United States and other countries to use the Montreal Protocol to regulate the potent greenhouse gases known as hydrofluorocarbons. Introduced to replace the ozone-destroying compounds outlawed by the Montreal treaty, these refrigerants are currently managed under the United Nations climate framework because of their greenhouse effects. Many argue

project, which has documented global land-use trends through more than 3.7 million images dating back to 1972. Landsat 8 will collect at least 400 images per day at several visible and near-infrared frequencies, covering the planet every 16 days. The USGS has accommodated 11 million downloads since 2008.

Eye in the sky

The US Geological Survey (USGS) says that data from its latest environmental satellite, Landsat 8, are now publicly available (at go.nature.com/u81wkx; shown, an image over Northwest Arctic Borough in Alaska). The US\$855-million spacecraft extends the world’s longest continuous Earth-observation



USGS

EVENTS

Accelerator shut

Following a radiation leak, no experiments will take place at the Japan Proton Accelerator Research Complex (J-PARC), in Ibaraki prefecture, until early next year, the facility’s director, Yujiro Ikeda, said on 10 June. The leak occurred at the Hadron Experimental Facility on 23 May, when a proton beam damaged a gold target, releasing material that exposed 34 workers to low-dose radiation. A malfunction of the beam extraction unit was blamed, and an

AXEL GRIESCH/MPFS

investigation is under way. All experiments up to the end of July are cancelled. J-PARC was already scheduled to close from August this year until late January 2014 for maintenance.

China in space

China launched its fifth crewed space mission on 11 June. The *Shenzhou 10* space capsule, carrying three astronauts, is scheduled to dock with the country's orbiting *Tiangong 1* space module on 13 June, and aims to return to Earth on 26 June. It will be the last mission to *Tiangong 1*. China plans to launch two more modules before 2016, in the run-up to building a crewed space station by 2020 (see *Nature* 473, 14–15; 2011).

PEOPLE

Max Planck chief

Martin Stratmann (pictured), a chemist at the Max Planck Institute for Iron Research in Düsseldorf, Germany, was elected on 6 June as the next president of the Max Planck Society in Munich, Germany's largest non-university basic-research organization. Stratmann, 59, will take office in June next year, replacing developmental biologist Peter Gruss, who has presided over the Max Planck Society since 2002. The society runs more than 80 research institutes in Germany, with an overall



budget this year of about €2 billion (US\$2.6 billion). See go.nature.com/1jccau for more.

RESEARCH

How to cut carbon

With carbon emissions climbing by 1.4% to 31.6 gigatonnes in 2012, the world is headed for a long-term temperature rise of 3.6–5.3 °C, said the International Energy Agency in a 10 June report. The agency, based in Paris, endorsed four cost-effective policies to help set the world on path to a 2 °C rise. Its major recommendation is the adoption of energy-efficiency measures such as performance standards for lighting, heating and road vehicles. Other policies involve limiting the construction and use of inefficient coal-fired power plants; minimizing methane leaks from the oil and gas industry; and phasing out fossil-fuel subsidies.

BUSINESS

Science networking

A professional networking site for researchers has raised US\$35 million from investors including Microsoft co-founder Bill Gates. ResearchGate, headquartered in Berlin, was founded in 2008 by two virologists and says it now has more than 2.9 million members. It is one of a number of sites (including Academia.edu and Mendeley) that aim to be hubs for scientists to connect and share publications. ResearchGate announced the latest funding on 4 June, but declined to disclose investment raised in two previous funding rounds.

Ups and downs

Pharmaceutical firm AstraZeneca will pay US\$560 million up front for Pearl Therapeutics in Redwood City, California, which is developing an inhaled treatment for lung conditions such as bronchitis and emphysema (collectively known as chronic obstructive pulmonary disease). London-based AstraZeneca announced the purchase (which could cost up to \$1.15 billion if other milestones are met) on 10 June. Six days earlier, it said that it had pulled the plug on a once-promising rheumatoid arthritis drug, fostamatinib, after a series of

COMING UP

17 JUNE

The European Commission, the European Parliament and the Council of the European Union negotiate the final details of their 2014–20 research-funding programme, Horizon 2020.

17–21 JUNE

The latest research into detecting nuclear explosions is presented at a conference hosted by the Comprehensive Nuclear-Test-Ban Treaty Organization in Vienna. go.nature.com/xtgtso

late-stage clinical-trial failures. The company returned licensing rights for the drug to biotechnology firm Rigel in South San Francisco, California, from whom it bought the rights for more than US\$100 million in 2010 (see *Nature* <http://doi.org/fcxh6c>; 2010).

Diabetes debate

Advisers to the US Food and Drug Administration say that the agency should ease its restrictions on access to Avandia (rosiglitazone), a diabetes drug linked to increased heart risk. Three years ago, US regulators sharply curtailed access to the drug, which is made by London-based GlaxoSmithKline. European authorities pulled the drug from the market altogether (see *Nature* 467, 505; 2010). But the advisory committee, meeting on 6 June, noted that a re-analysis of a pivotal clinical trial suggested that people taking the drug were not more likely than others to die from heart complications. See go.nature.com/4zlswa for more.

► **NATURE.COM**

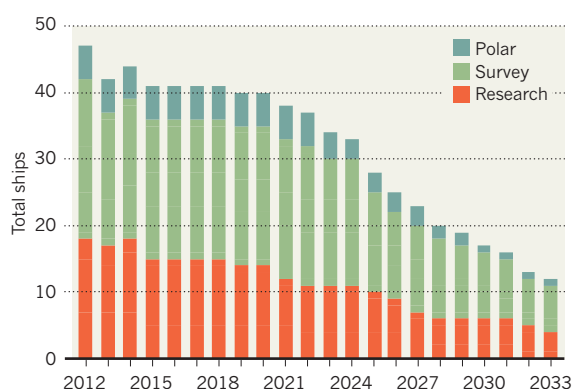
For daily news updates see: www.nature.com/news

TREND WATCH

A government report on the status of the US oceanographic fleet says that fuel costs for research ships have increased by 400% since 2003, and that pressure on budgets has led to some US vessels being disposed of or laid up. As vessels are retired, the fleet is set to shrink rapidly unless new ships are built or old ones overhauled (see chart). Even new vessels currently being built will not stem the fleet's decline if they enter service as scheduled. See go.nature.com/fvcs5z for more.

US RESEARCH FLEET NEEDS RESUPPLY

The US oceanographic fleet is on course to roughly halve by 2026 as ships are retired from service.



SOURCE: NATL OCEAN COUNCIL

NEWS IN FOCUS

CANCER Clinical trials could soon bow to 'master' protocol **p.146**

MICROBIOLOGY Regulators take on messy field of faeces transplants **p.147**

ZOOLOGY Usain Bolt has nothing on a cheetah's acceleration **p.150**



PRIZES Upstart science mega-awards challenge old-school Nobels **p.152**

STR/AFP/GETTY



China's decision to reduce its carbon dioxide emissions should help to alleviate the dangerously high levels of air pollution in the country.

CLIMATE CHANGE

China gets tough on carbon

Cap-and-trade pilot schemes set stage for nationwide roll-out.

BY JANE QIU

China, responsible for about one-quarter of the world's carbon dioxide emissions, has ambitious goals to reduce them — but has been unwilling to set absolute targets for fear of slowing economic growth. There are now signs that its position is changing.

On 18 June, the country will launch an emissions-trading scheme in the southern city of Shenzhen, marking its first attempt to cut emissions using market mechanisms. Under the scheme, more than 630 industrial and construction companies will be given quotas for how much carbon dioxide they can emit. Companies that pollute more than they are allowed will have to buy credits from cleaner counterparts that reduce emissions below their quota — thereby creating a price for the greenhouse gas.

Another six such cap-and-trade schemes will

be rolled out by the end of the year in the cities of Beijing, Tianjin, Shanghai and Chongqing, and the provinces of Guangdong and Hubei. The trial will cover 864 million tonnes of carbon dioxide by 2015 — around 7% of China's total emissions and about the total amount emitted by Germany each year, according to a report by the London-based analyst firm Bloomberg New Energy Finance. These regional pilot schemes will set the stage for the nationwide carbon market that is scheduled to launch in 2016.

China has committed to cutting its carbon intensity — carbon emissions per unit of gross domestic product — by 40–45% of 2005 levels by 2020, which allows for increases in emissions, although at a slower rate. The initial emissions limits for the regional schemes will be set by

applying the carbon-intensity targets to the emissions of individual companies. In 2016, this system will be scaled up nationally, again in line with carbon-intensity targets.

After 2020, this plan is likely to be replaced with an absolute cap that would require a decline in overall emissions covered under the scheme. Such a move will depend on the effectiveness of an array of planned energy policies, researchers say. "It's not difficult from a technical point of view," says Xiang Gao, a member of China's climate-talks delegation and a researcher at China's National Development and Reform Committee (NDRC), the powerful ministry responsible for planning the country's economic and social development. "It's a matter of political will — which, in turn, will depend on whether the top leadership can be convinced that such a move is best for the country's economy and social stability," Gao says. ▶

➤ **NATURE.COM**
For more on China's emissions and pollution, see: go.nature.com/clipse

► Researchers say that China has reasons beyond climate change to implement emission caps. In the past few years, rampant air pollution has caused increased public resentment and social unrest across the country. “China may not have a choice any more,” says Knut Alfsen, head of research at the Centre for International Climate and Environmental Research in Oslo. “It’s just much better to control total emissions.”

A commitment from China to cap emissions “would breathe new life into climate talks”, adds Alfsen, who is also a member of the China Council for International Cooperation on Environment and Development, an international think tank that works closely with China’s cabinet and the NDRC. At the next climate-change summit, in Paris in 2015, nearly 200 countries will aim to reach a legally binding global agreement on emissions cuts, which would take effect in 2020. Kelly Sims Gallagher, an expert on energy and environmental policy at Tufts University in Medford, Massachusetts, says that an ambitious emissions cap from China “would send a strong political signal to the world” and would make it easier to

pass more aggressive climate legislation in the United States, where there is strong political resistance to national climate regulations.

Most researchers contacted by *Nature* are only cautiously optimistic that China can cap its emissions. A carbon ceiling for China “depends in part on how successful the pilot schemes will be”, says Lei Ming, an environmental economist at Peking University in Beijing. “We will have to cross the river by feeling the stones,” he says, citing the famous one-liner by the late reformist leader Deng Xiaoping.

One of the main challenges for the nationwide cap-and-trade scheme will be establishing its credibility. Verifying emissions, for instance, will be difficult in such a large country, says Gallagher. David Yuetan Tang, board secretary of the Tianjin Climate Exchange, which is in charge of one of the seven pilot emission-trading schemes, says that there is an institutional void about who will do this — and also a legal

void about how companies will be punished for fraudulent claims or emissions excesses. “This is absolutely paramount, because emission quotas are money,” he adds.

Moreover, whether emissions trading can work under China’s political system remains to be seen, critics say. “The energy market in China is not entirely free and has a lot of government interference and monopoly,” says Qi Ye, an environmental-policy researcher at Tsinghua University and director of the Beijing office of the international think tank Climate Policy Initiative. The price of electricity, for instance, is heavily controlled, he says, which could seriously diminish the impact of imposing a carbon price on electricity producers.

Emissions trading is just one of a series of energy and pollution policies due to be introduced in the next few years. For instance, Beijing is considering implementing a carbon tax to rein in pollution by sectors not covered by cap and trade, and continues to invest aggressively in renewable energy. It has also pledged to reduce the production and use of hydrofluorocarbons, powerful greenhouse gases used in refrigeration and air conditioning. ■

PERSONALIZED MEDICINE

‘Master protocol’ aims to revamp cancer trials

Pilot project will bring drug companies together to test targeted lung-cancer therapies.

BY HEIDI LEDFORD

In the push to match medical therapies to the genetic underpinnings of disease, lung-cancer treatments have been at the frontier. But the 1.6 million people diagnosed with this cancer every year will take scant comfort in knowing that of the past 20 late-stage trials of drugs to treat it, only two yielded positive results. And in only one of those 20 were patients chosen systematically by screening for biomarkers such as relevant blood proteins or DNA sequences.

Now, an ambitious project aims to improve those success rates and speed new treatments to market by matching companies with the patients whose tumours are most genetically relevant to the therapies they are trying to develop. The project is slated to launch next year and, if successful, could be expanded to other cancers.

The project was spearheaded by the Friends of Cancer Research, a think tank and advocacy group in Washington DC, and has won the support of the US National Cancer Institute and the US Food and Drug Administration (FDA). The

idea is to streamline the drug-approval process by bringing pharmaceutical companies together to test multiple experimental drugs in late-stage clinical trials under a single, ‘master’ protocol. “The drive is to make the whole process of personalized medicine more efficient,” says Eric Rubin, vice-president of oncology clinical research at Merck, a pharmaceutical firm based in Whitehouse Station, New Jersey.

PLUG AND PLAY

Launching a large, late-stage clinical trial typically takes more than two years and requires some three dozen administrative and regulatory approvals. To simplify this tangle, the master protocol will create an experimental plan to test several candidate drugs in hundreds of clinics across the United States. The initial protocol is expected to include up to six drugs; others may be added later, without the need for fresh protocol approval each time. “It’s like a Plug and Play,” says David Gandara, an oncologist at the University of California, Davis, who is in charge of drafting the plan. “So you don’t waste time over and over.”

Gandara has advocated this approach for the past decade, but the FDA and the pharmaceutical industry voiced support only recently — swayed by a growing body of data revealing that cancers are, in effect, many rare diseases with different genetic roots (see *Nature* 455, 148; 2008). A genetically targeted drug may work, but only in a fraction of cases. Such rare effects could easily be overlooked in a trial that contains a mix of patients whose cancers have heterogeneous causes, and the costs for drug companies to sort them all and run scores of separate trials are prohibitive.

Under the master protocol, by contrast, patients will be screened for various biomarkers and assigned to trials for drugs that are most likely to be effective. The approach does away with the need for patients to undergo multiple screenings: participating companies could enrol them from a large, central pool. It also eases pressure on the (often minute) tissue samples taken during lung biopsies, because many tests can be done at the same time, says Rubin.

A similar model is already being tested in two smaller clinical trials for breast and lung

cancers (see *Nature* **464**, 1258; 2010). Both trials involve multiple biomarkers, drugs and clinics, and both won support from pharmaceutical companies. But that does not mean that drug companies will embrace a larger, more developed venture, says Roy Herbst, an oncologist at the Yale School of Medicine in New Haven, Connecticut, who chairs the steering committee of the master-protocol project. It is much easier to coax a company into a group effort for a small, early trial than to persuade it to give up any measure of control over a late-stage one crucial for gaining regulatory approval.

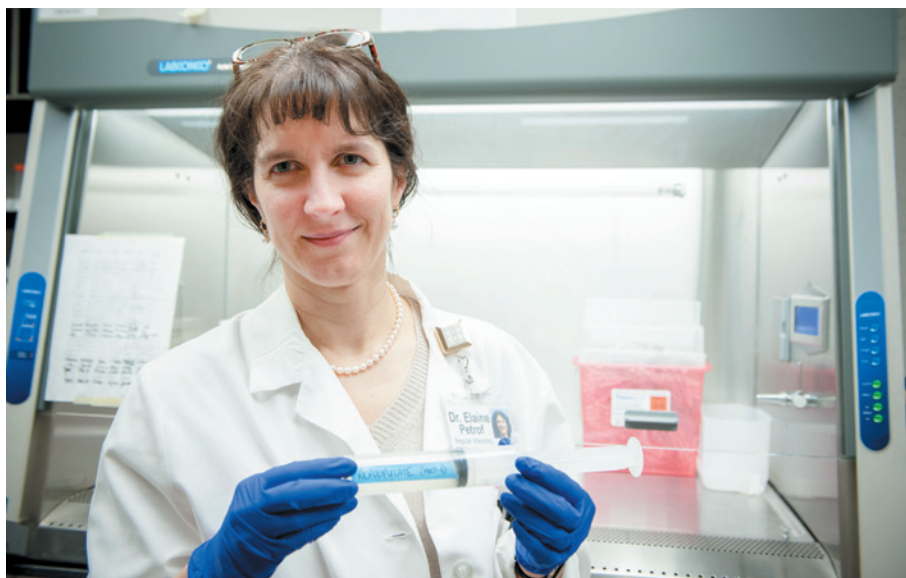
Companies also prefer to maintain control of proprietary information rather than deposit early results into centralized databases. "It's a challenge," says Herbst. "Many of them might think they can do it alone, and may worry about losing autonomy."

The project's organizers tried to address industry concerns early on, says Ellen Sigal, founder and chairwoman of Friends of Cancer Research. At a planning meeting in March, representatives from more than 20 drug companies were assured that the FDA supports the protocol and has statisticians working to help shape it — making the agency more likely to feel comfortable basing approval decisions on data from the trial. Organizers also pledged to have a neutral third party monitor the trial, to ensure that drugs made by competing companies would not be directly compared.

Gandara hopes that the speed and lower costs will also draw industry partners. Late-stage clinical trials can cost between US\$50 million and \$100 million; Gandara estimates that the master protocol could cut that to \$25 million or less.

Companies might also be wooed by easy access to the National Cancer Institute's vast network of treatment centres and clinicians who are experienced in conducting clinical trials. That network will allow the trial to be conducted at 500 sites in the United States and Canada and enable it to enrol up to 1,000 patients a year.

Thus far, the downside of participating seems minimal, says Richard Gaynor, head of oncology-product development at Eli Lilly, a pharmaceutical firm based in Indianapolis, Indiana. "It will be an interesting experiment," he says. ■



Elaine Petrof has invented a synthetic stool that could reset a patient's gut bacteria to cure infections.

GASTROENTEROLOGY

FDA gets to grips with faeces

Regulator triggers efforts to standardize faecal transplants.

BY BETH MOLE

The brown slurry is piped through tubes into the top of the human body — or the bottom. It can even come in pill form. For years, doctors have been transferring faeces into ill people's intestines to replace resident microbes with a fresh batch. The procedure is often a therapeutic success, but protocols for it vary wildly. As it steadily grows more popular, regulators are now working to define what a standard faecal transplant should be, and how to deliver one safely.

During a public workshop last month at the US National Institutes of Health in Bethesda, Maryland, the Food and Drug Administration (FDA) reaffirmed that it has authority over faecal transplants. The agency had said this

for years to researchers and companies who asked privately, but the workshop was the first public forum in which the FDA broadcast that it regulates faeces like a drug.

Clinical trials of the procedures are not affected, because they were already subject to approvals from the agency. But US doctors performing faecal transplants as treatments must now submit an Investigative New Drug application to the FDA with details about their protocols. (The agency then has 30 days in which it can intercede and stop an experiment.) Jay Slater, director of the division of bacterial, parasitic and allergenic products at the FDA in Silver Spring, Maryland, says that the move is a crucial way for the agency to make sure that protocols are safe. But he adds that the FDA wants to avoid being too prescriptive for ►


**MORE
ONLINE**

Q&A



Paul Nurse talks about the science plans of London's new Francis Crick Institute go.nature.com/bpgdnz

MORE NEWS

- Dust grains point to meteorite origin for Tunguska blast go.nature.com/5h7wbq
- Skeletons reveal that the Medicis had rickets go.nature.com/yqiwyh
- Mouse model of OCD triggered and treated with optogenetics go.nature.com/yy6pli

NATURE PODCAST



What lives beneath the sea floor, new science prizes, and the resurgence of nature writing nature.com/nature/podcast

GUT INSTINCT

Faecal transplants are an increasingly popular way to treat infections of *Clostridium difficile*, but approaches vary wildly.

Clinic	Route	Stool amount	Stool freshness	Blending method	Patients treated	Claimed success rate
Mayo Clinic, Rochester, Minnesota	Colonoscopy	50 g	< 6 hours old	Lab paddle blender	~40	90–95%
Nebraska Medical Center, Omaha	Nasal tube	30–50 g	< 6 hours old	Blender	17	94%
The Bright Medicine Clinic naturopathic practice, Portland, Oregon	Enema	50–300 g	Frozen or < 6 hours old	Blender	8	88%
Kingston General Hospital, Canada (clinical trial)	Colonoscopy	100 ml (synthetic)	Cultured	Hand-mixed	~30	Planned trial
Thomas Louie's private practice, Calgary, Canada	Capsules	0.47 ml per pill	< 6 hours old	Food mill	33	100%

► now, so that it can adopt the most effective, advanced protocols as they are developed.

Although it may be years before the agency weighs in on which method is the safest, it has ignited a debate among researchers over how faeces should be screened, processed, delivered — or even synthesized.

With faecal transplants, doctors aim to reestablish healthy microbe populations in the guts of patients. The procedure seems especially effective for people infected with *Clostridium difficile*, a diarrhoea-causing bacterium that in the past two decades has become more prevalent and antibiotic-resistant in the United States, where it now kills an estimated 14,000 people each year. A 2011 review of data from more than 300 patients concluded that faecal transplants can cure 92% of people with recurring *C. difficile* infections for which antibiotics prove ineffective (E. Gough *et al. Clin. Infect. Dis.* 53, 994–1002; 2011).

But there are many issues and unanswered questions. The method's success against *C. difficile* has led to an "outrageous exuberance", says Amee Manges, an epidemiologist at the University of British Columbia in Vancouver, Canada, who led the review. Some doctors are using faecal transplants to treat other conditions, for which effectiveness is less established — or not established at all. Faeces, if not properly screened, can transmit disease. Furthermore, it is too early to know which of the many protocols is the most effective. "Everybody has their preferences," says Manges (see 'Gut instinct'). Resourceful individuals can even get in on the act at home, by following step-by-step enema instructions from online videos.

The wide variation in clinical practice starts at the very source of the 'drug'. Although evidence is lacking, some researchers suspect that the best stool comes from a patient's blood relatives, who have genetic and environmental similarities with the patient that might influence their gut microbes. Other doctors use anonymous donors.

Preparation methods also differ. Some researchers freeze the stool for convenience, to

use later. Others insist that it must be fresh — 6 hours old or less — to ensure that the bacteria do not die or change their behaviour during their time outside the colon. Fresh or frozen, the stool is mixed with a liquid — usually saline, although some researchers have tried water or even milk. Others are exploring synthesizing faeces from scratch (see 'How to make a stool').

When it comes to the mode of delivery, some

FAKE FAECES

How to make a stool

The runny, cloudy substance, developed by Elaine Petrof, an infectious-disease researcher at Queen's University in Kingston, Canada, is one of the first prototypes of synthetic stool: a mixture of 33 microbes individually isolated from the faeces of a healthy donor and then recombined.

Petrof chose the donor after a stringent screening protocol, selecting a woman who was infection- and parasite-free and clear of chronic diseases and drugs, and who had taken antibiotics only once in her life, long ago.

For the microbe mix, Petrof chose a combination of beneficial bacteria and others known to support the overall microbial community, and threw the rest away. The chosen bacteria were grown in pure culture, then mixed together in saline solution at ratios that replicated their original proportions in the stool.

In January, Petrof and her colleagues published results showing that the slurry of microbes, called Re-POOPulate, could be used in faecal transplants, curing two people of life-threatening *Clostridium difficile* infections (E. O. Petrof *et al. Microbiome* 1, 3; 2013). To confirm the results, the team plans to enrol 30 people in a clinical trial. **B.M.**

researchers use enemas — easy to administer, but good for transplanting stool into only the lower end of the colon. Others use colonoscopies, which are more invasive but ensure that the stool makes it farther into the intestines.

Johan Bakken, an infectious-disease consultant at the University of Minnesota Medical School in Duluth, who has used faecal transplants to treat 120–130 people with *C. difficile* infections, delivers the preparation in a tube threaded through a patient's nose into the small intestine. This, he argues, may be safer for people with disease-weakened intestinal walls that could be torn during enemas or colonoscopies. Thomas Louie at the University of Calgary in Canada avoids tubes altogether: so far, he has treated 33 people by delivering stool microbes orally, wrapped in three layers of medical-grade gelatin capsules.

Researchers generally agree that donors must be screened using standardized procedures if faecal-transplant therapies are to flourish. Many are concerned that inadequately checked material could contain pathogens, just as blood transfusions sometimes caused transmission of hepatitis C in the days before screening. Faecal screens tend to include tests for blood-borne pathogens such as HIV and hepatitis viruses, as well as intestinal pathogens and parasites. But some scientists have collected anecdotes of donors who were not even tested for obvious pathogens such as HIV and *C. difficile*.

One of the most pressing questions is which diseases can be treated effectively with faecal transplants. In addition to *C. difficile* infections, researchers have used the procedures to treat chronic problems such as Crohn's disease, inflammatory bowel disease and multiple sclerosis, but in very small case studies. Some clinics are even recommending faecal transplants for obesity, Parkinson's disease or autism spectrum disorder — although most doctors remain sceptical. More data and oversight are needed to enable researchers to learn which applications work and which do not, says Gary Wu, a gastroenterologist at the University of Pennsylvania in Philadelphia. "Stool is a very complex mixture that we don't fully understand," he says.

Wu expects that researchers will eventually move to synthetic stool, a potentially safer and more consistent concoction. "But right now, we're not at that point," he says.

Some researchers expressed frustration with the FDA's move to regulate faecal transplants. Infectious-disease specialist Trevor Van Schooneveld of the University of Nebraska Medical Center in Omaha, has performed about 20 such transplants since 2011, working with gastroenterologists. But in the past few weeks, he has turned away three patients while he submits the required Investigative New Drug application. Van Schooneveld questions whether the agency should preside over an organic, personal substance, rather than a drug. "How the FDA plans to regulate human faeces is a mystery to me," he says. ■

PHOTONICS

Light flips transistor switch

Photons emerge as competitors to electrons in computer circuits.

BY DEVIN POWELL

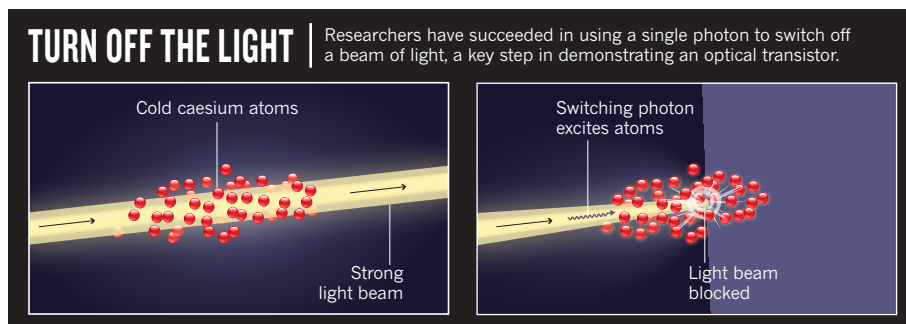
Transistors, the tiny switches that flip on and off inside computer chips, have long been the domain of electricity. But scientists are beginning to develop chip components that run on light. Last week, in a remarkable achievement, a team led by researchers at the Massachusetts Institute of Technology (MIT) in Cambridge reported building a transistor that is switched by a single photon.

Conventionally, photons are used only to deliver information, racing along fibre-optic cables with unparalleled speed. The first commercial silicon chip to include optical elements, announced last December, did little to challenge the status quo. The on-board beams of light in the device, developed at IBM's research centre in Yorktown Heights, New York, merely shuttle data between computer chips.

Now, Wenlan Chen of MIT and her colleagues have taught light some new tricks, using a cloud of chilled caesium atoms suspended between two mirrors. Their transistor is set to 'on' by default, allowing a beam of light to sail through the transparent caesium cloud unmolested. But sending in a single 'gate' photon turns the switch off, thanks to an effect called electromagnetically induced transparency. The injected photon excites the caesium atoms, rendering them reflective to light trying to cross the cloud (see 'Turn off the light'). One photon can thus block the passage of about 400 other photons, says Chen, who presented the result on 7 June at a meeting of the American Physical Society's Division of Atomic, Molecular and Optical Physics in Quebec City, Canada.

The ability to turn a strong signal on and off using a weak one fulfils a key requirement of an optical transistor. "Nothing even came close before," says physicist Ataç Imamoglu of the Swiss Federal Institute of Technology Zurich, who called the experiment "a true breakthrough". In theory, the hundreds of photons, controlled by the triggering photon, could fan out and switch off hundreds of other transistors in an optical circuit.

With its exotic clouds of atoms and bulky equipment, the proof-of-principle transistor is unlikely to become a component in everyday computers. But it could be a useful tool for studying how photons interact at the quantum level — potentially leading to a quantum transistor that flips, not a one or a zero as in



classical computing, but a fuzzy bit of quantum information.

A more practical optical transistor debuted in April 2012 at Purdue University in West Lafayette, Indiana, where electrical engineer Minghao Qi has made one that is compatible with the semiconductor industry's existing manufacturing techniques¹. "The advantage of our device is that we have it on a silicon chip," says Qi.

In this case, the beam of light to be switched on and off enters and exits along a channel, etched in the silicon, that sits next to a parallel channel. In between the two rails is an etched ring. When a weaker light beam courses through the second optical line, the ring heats up and swells, interfering with the main beam and switching off the transistor. This switch can flip on and off up to 10 billion times per second.

And the output beam can fan out and drive two other transistors, meeting one of the established requirements² for an optical transistor set out in 2010 by David Miller, a physicist at Stanford University in California. Other criteria include matching the frequency of the exiting signal to the input frequency and keeping the output clean, with no degradation that could cause errors. "Making an optical transistor that really satisfies the necessary criteria is very hard," says Miller.

Still, Qi does not expect to challenge the electronic transistor with his optical analogue, which consumes a lot more power and runs much more slowly. "We want to complement the Intel transistor," he says. "We don't want to replace it." He hopes to find a foothold in niche

markets, such as equipment for scrambling cable channels and military technologies that could benefit from light's imperviousness to an electromagnetic attack.

Routers that guide information through the Internet could also be amenable to optical transistors and switches. At present, these stopping points in the network convert optical signals travelling through fibre-optic cables into electrical signals; these are then processed, converted back to light and sent on their way. A router in which one beam of light pushes another in the appropriate direction — with no conversions involved — could in principle be faster and consume less energy.

A popular candidate for such switches are quantum dots, small semiconductor crystals that behave like atoms. In one particularly sensitive quantum-dot switch, a beam of light is first guided along a material dotted with holes, called a photonic crystal. The light can pass through a quantum dot placed in its path without changing course. But if a pulse of light is sent in just ahead of that beam, it can induce an interaction between the dot and the crystal that scatters the beam and sends it on a different path.

Reported in May 2012 by Edo Waks of the Joint Quantum Institute at the University of Maryland in College Park and colleagues³, it switches when struck by a pulse of 140 photons. In principle, that is a small enough amount of energy to rival conventional routers.

But the switch still faces a practical obstacle common to all of these emerging optical technologies. The lasers that supply the devices with light consume considerable energy, offsetting any savings. "Right now," says Waks, "the overhead is what's killing us." ■

➔ **NATURE.COM**
For more on
alternative
transistors, see:
go.nature.com/r6dl2q

much more slowly. "We want to complement the Intel transistor," he says. "We don't want to replace it." He hopes to find a foothold in niche

1. Varghese, L. T. et al. Preprint at <http://arxiv.org/abs/1204.5515> (2012).
2. Miller, D. A. B. *Nature Photon.* **4**, 3–5 (2010).
3. Bose, R., Sridharan, D., Kim, H., Solomon, G. S. & Waks, E. *Phys. Rev. Lett.* **108**, 227402 (2012).



KRYSTY GOLABEK

Cheetahs can hunt successfully in open environments and in dense vegetation.

ZOOLOGY

Speed test for wild cheetahs

State-of-the-art collar reveals animal's quick reflexes and phenomenal acceleration.

BY MATT KAPLAN

The cheetah crouches in the undergrowth. When a young antelope strays a little too far from the herd, the cheetah explodes out of the bush — and, with a burst of speed unrivalled in the natural world, brings down its next meal.

Or so we have assumed. But the first study to collect data on the animal's movements in the wild reveals that, contrary to popular opinion, a cheetah's sheer speed is not its only weapon when it comes to hunting. Its success as a predator also hinges on its lightning reflexes and its ability to accelerate faster than a Ferrari.

Determining just how fast animals run is no easy task. In zoos, captive cheetahs lured to run in a straight line can attain speeds of up to 29 metres per second — nearly 105 kilometres per hour (N. C. C. Sharp *J. Zool.* **241**, 493–494; 1997), more than double the top speed achieved by a human sprinter. But nobody had been able to determine whether the animals actually reach these speeds in the wild.

Armed with lightweight solar-powered collars fitted with Global Positioning System and inertial-measurement technologies, a team led by Alan Wilson at London's Royal Veterinary College has precisely tracked wild cheetahs during hunting (see page 185).

The team first tested the accuracy of their

collars on a dog by letting the animal loose on a beach; that way, the information collected could be cross-referenced with the paw prints left in the sand. The collars proved to be astonishingly precise, down to a level of 0.2 metres.

The researchers then travelled to the Okavango Delta region of northern Botswana, where they used darts to sedate and collar five cheetahs. The 17 months spent collecting the data was the easy bit; what posed the challenge was making sense of the data, which confounded expectations.

Out of a total of 367 runs, the fastest speeds achieved by the five individuals were 25.9, 25.4, 22.0, 21.1 and 20.1 metres per second — all well short of the record set by captive cheetahs. Moreover, most hunts involved only moderate speeds, with the average top speed coming in at 14.9 metres per second. But although the wild cheetahs did not run as fast as their captive kin, they demonstrated other athletic abilities that researchers had not been able to measure before.

The data revealed the cheetahs' acceleration powers to be up to 120 watts per kilogram — about double the power of the swiftest greyhounds and more than four times that of Usain Bolt during his record-breaking 100-metre

sprint in 2009. The cheetahs were also able to slow down rapidly, absorbing energy at a rate up to three times greater than that known for the best-performing polo horses — animals that are bred to be agile.

When the researchers combined this information with field observations of cheetah kills and terrain information provided by Google Earth, they realized that the cheetahs were often hunting successfully amid thick vegetation by making sharp turns and sudden stops. "We have always thought of cheetahs as sprinters, but now it looks as though sprinting is only part of the story," says Wilson.

"It is remarkable," says evolutionary biologist David Carrier at the University of Utah. "Both agility and manoeuvrability turn out to be at least as important to these animals as speed."

Anticipation of what Wilson's collars might reveal in the future is growing fast. "I really wonder if cheetahs living on the open savannahs will yield the same sorts of results," says Jack Grisham, coordinator of the cheetah species survival plan for the Association of Zoos and Aquariums.

Carrier, meanwhile, hopes the collars will soon be used to study the movements of other animals in the wild. "Simultaneous recordings from each member of a pride of lions or a pack of wild dogs would prove fascinating," he says. ■

NATURE.COM
For video footage
of a cheetah giving
chase, see:
go.nature.com/adyfnt

FUNDING

UK scientists fear further cuts

Funding jitters rife ahead of government spending review.

BY DANIEL CRESSEY

With anxiety rising about what the immediate future may hold for Britain's science funding, the man responsible for the nation's finances is trying to allay researchers' fears.

Science "is a personal priority for me", chancellor of the exchequer George Osborne told reporters on 6 June after a ceremony to mark the completion of the roof of the new £650-million (US\$1.1-billion) Francis Crick Institute under construction in London.

On 26 June, Osborne is set to unveil the next comprehensive spending review (CSR), which sets spending for government departments. He said that he hoped to make clear the government's "long-term commitment" to research in the new review, but scientists fear another budget freeze. Asked if he could cut science after his supportive statements, the chancellor said that he would not pre-empt the CSR but added: "You can read between the lines that I'm going to do everything I can to make sure Britain has a bright scientific future."

The previous spending review, in 2010, set budgets for government departments for the four financial years to 2014–15. June's CSR will apply to just the 2015–16 fiscal year — because a new budget will be crafted after a general election in May 2015 — and Osborne has made it clear he wants cuts to most departments.

Despite its short duration, this CSR is important, says Kieron Flanagan, a science-policy researcher at Manchester Business School. "You can do damage in one year" if spending is cut back severely, and whoever wins the election in 2015 would be likely to work from the existing framework, he says.

Analysts are especially keen to know what the government will do with the 'ring fence' that was placed around the science budget in 2010, freezing it at £4.6 billion a year. The fence spared core spending areas — such as grants that are awarded by the country's research councils — from the cuts inflicted on other public sectors, although the science budget still lost money in real terms each year. The umbrella group Universities UK has calculated that, when inflation is taken into account, the deficit is £600 million over the current four-year CSR period.

And, in any event, the ring fence had holes. The 2010 CSR moved capital spending in

science — monies allotted to large infrastructure projects such as buildings and facilities — outside the ring fence, away from the core science budget. That made infrastructure vulnerable to cuts, and projects such as the United Kingdom Infrared Telescope in Hawaii face closure as a result (see *Nature* **486**, 168; 2012).

Many policy analysts expect the ring fence around science funding to be retained in the new CSR. But some worry that it may

"The absolutely crucial thing is we fund basic scientific research — including in medicine."

be removed or that additional categories of science money could be moved outside it.

One rumour in circulation is that the Medical Research Council (MRC), which is a major funder of UK medical research, will be moved from the Department for Business, Innovation and Skills — the department in charge of the science budget — to the Department of Health, where it might be more vulnerable to cuts or to a change in research focus. In a 6 June statement, Ted Bianco, acting director of the biomedical-funding charity the Wellcome Trust, called the prospect "ill-advised and potentially damaging", adding that it would shift the balance "from fundamental to applied research when both are essential to medical progress".

Osborne would not comment during the Crick Institute event on a move for the MRC, but said that "the absolutely crucial thing is we fund basic scientific research — including basic scientific research in medicine — and I'm not prepared to do anything that puts that at risk".

James Wilsdon, a science-policy researcher at the University of Sussex, says that another year of flat cash for science would be "painful but survivable". Deeper cuts, he says, would be another matter. ■

CORRECTION

The News Feature 'The gun fighter' (*Nature* **496**, 412–415; 2013) wrongly implied that blogger David Codrea had 'outed' gun researcher Garen Wintemute. Wintemute had in fact publicized his own work before Codrea's 2007 blog post.



As reactions to winning a multimillion-dollar prize go, Alexander Polyakov's words were less than gushing. It was the culmination of a glittering award ceremony in Geneva, Switzerland, in March, hosted by Hollywood actor Morgan Freeman and featuring an operatic interlude from British singer Sarah Brightman. After a hushed pause, the physicist from Princeton University, New Jersey, was revealed as the winner of the 2013 Fundamental Physics Prize and an accompanying payment of US\$3 million. "This new prize is an interesting experiment," a flustered Polyakov said moments later, after walking off stage clutching his sculptured silver trophy. "Such big prizes could become very influential and

they can have a positive impact, or they can be very dangerous."

Polyakov's ambivalence echoes the sentiments of many scientists towards the rash of big-money science prizes that have emerged over the past year. Sponsored by billionaire entrepreneurs, including Russian Internet mogul Yuri Milner, Facebook supremo Mark Zuckerberg, Google co-founder Sergey Brin and property developer Samuel Yin, the new awards outstrip the \$1.2-million Nobel prizes in monetary value. In addition to Milner's Fundamental Physics Prize, the Internet billionaires have together created the Breakthrough Prize in Life Sciences, and Yin has introduced the Tang Prize as an Asian complement to the Nobels.

The founders of these 'new Nobels' hope that

the winners will act as role models, inspiring future generations to pursue science, and that they will attract status and funding to the entire discipline. "We wanted to choose an amount that would be shocking," says Anne Wojcicki, a biotechnology analyst and Brin's wife, who sits on the board of the Breakthrough Prize. "We wanted to create science superheroes."

But the lavishness and ambition of the prizes have sparked criticism. "I don't want to run these awards down, but I find it offensive that people are trying to either borrow the prestige of the Nobel, or buy it," says Frank Wilczek, a physicist at the Massachusetts Institute of Technology in Cambridge, who won a share of the Nobel Prize in Physics in 2004. "The suspicion is that these provide more benefit

PAUL JACKMAN/NATURE

to the egos of the founders than to science,” adds Jack Stilgoe, a lecturer in science policy at University College London.

And although they support the goals of the prizes, critics say that the strategy for achieving them is at best misguided, and at worst, could backfire. By bestowing riches on a few individuals, they say, the prizes could funnel money and attention towards people and fields that are already prestigious and well funded or, in some cases, could reward weak scientists or untested ideas. “Prizes are a good thing, but the question is, if your goal is to help science, are large prizes the most efficient way to do that?” asks Wilczek.

TOWARDS THE NEXT GENERATION

First awarded in 1901, the Nobel prizes have become established as the benchmark of excellence in the sciences. Since then, other awards have sprung up and gained prestige within specific disciplines. Some, such as the Fields Medal and the Abel Prize for mathematics, were designed to reward achievements in disciplines that are not covered by the Nobels. Others, such as the Lasker Awards for medical sciences, have gained a reputation for predicting future Nobel winners.

The Fundamental Physics Prize was the first of the latest breed of awards (see ‘Follow the money’). It burst on the scene in July 2012, when Milner announced that he had given nine awards of \$3 million each to prominent theoretical physicists, and that he planned to sponsor one additional award each year. (Polyakov was the first single-award winner.) Milner, who himself pursued graduate studies in theoretical physics, says that he wants to show that foundational research can be as financially rewarding as careers in sports, entertainment or business; indeed, he chose the size of the prize to mirror the type of annual earnings seen in the financial world. “The best minds should make at least as much as any trader on Wall Street,” Milner says.

In January, Yin launched the Tang Prize, four awards of 40 million Taiwanese dollars (US\$1.3 million) for the winners, plus a grant of 10 million Taiwanese dollars each for their research. The Tang Prize will be awarded every two years from 2014 onwards and will recognize advances in sustainable development, biopharmaceutical science, Chinese studies and law. “For the past 100 years, it is mainly Western countries and Western research institutions that have fostered talented researchers,” Yin says. “Now, as Asian economies are developing well, we should also shoulder part of the responsibility and contribute to world development.”

The Breakthrough Prize in Life Sciences,

announced in February, also originates with Milner — but this time, he brought in friends and colleagues including Zuckerberg, Wojcicki and Brin. They split \$33 million equally between each of 11 laureates, and have committed to giving five new prizes of \$3 million each year. “We all have a background in science, though we weren’t all the best students,” says Wojcicki. “This is a way for us to reconnect with that and to give something back.” March saw the inaugural award of the United Kingdom’s £1-million (US\$1.5-million) Queen Elizabeth Prize for Engineering, which is supported by charitable



Alexander Polyakov receives the 2013 Fundamental Physics Prize from actor Morgan Freeman at a ceremony modelled on the Oscars.

donations from corporate sponsors and was set up by the Royal Academy of Engineering explicitly to give engineers a taste of the glamour and recognition that comes with the Nobels.

After the initial surprise at the big sums involved, the first question in most people’s minds was how the winners would spend their cash. “I really admire these billionaires for wanting to give back to science — but I do hope some of these large amounts goes into research,” says geneticist and entrepreneur Craig Venter, of the J. Craig Venter Institute in San Diego, California. “It’s not so great if the winners just go and buy a bigger house with their prize money.”

Many of the winners seem a little sheepish about their windfall; those who are willing to be interviewed tend to mumble that they have not yet decided what proportion to keep and what to give to research. One person who has decided is Tejinder Virdee, a particle physicist at Imperial College London, who in December 2012 shared a \$3 million ‘special’ Fundamental Physics Prize with six other leaders of the hunt for the Higgs boson at the Large Hadron Collider (LHC) at CERN, Europe’s particle-physics lab near Geneva. (The special prizes were in addition to the nine prizes Milner had previously announced.) Virdee plans to use his winnings to pay for science equipment in

schools in sub-Saharan Africa and to support an exchange programme to bring teachers from these schools to visit the LHC. “I wanted to find the way to get the most leverage out of the money,” he says. “It costs relatively little to train a teacher, but in turn they could reach 500 students in the next few years.”

Even if winners invest in their work, some researchers worry that the prize money will largely reward those who already have ample funding and recognition. Although the physics prizewinners do not include any Nobel laureates, between them they have won pretty much every other major award, including the Wolf Prize in Physics, the Fields Medal for mathematics and the MacArthur ‘genius grant’. “These are not underfunded or unappreciated people,” says Peter Woit, a mathematician at Columbia University in New York. Many of the Breakthrough Prize winners are already regarded as shoo-ins for future Nobels; one, Shinya Yamanaka, a stem-cell biologist at Kyoto University in Japan, already won a share of a Nobel last year.

This means that the prizes could end up increasing the divide between the scientific haves and have-nots. “There’s a huge disparity between the money that the big-wigs are getting and that going to other scientists,” says Bob O’Hara, a biostatistician at the Biodiversity and Climate Research Centre in Frankfurt, Germany. “US\$33 million could fund my whole institute for three to four years.”

O’Hara and other researchers also complain that the Breakthrough Prize recognizes the same popular fields as the Nobels. “One frustration for biologists is that the Nobel prize is focused on physiology and medicine and so it neglects other areas of the life sciences, which are just as important,” O’Hara says. The Breakthrough Prize does little to redress the balance, overlooking areas such as ecology and evolutionary biology in favour of research into molecular biology and disease. What is more, O’Hara says, “there was an emphasis on diseases of the rich, in the West, at the expense of diseases that are prevalent in the developing world”.

Wojcicki notes that there is a catch-22: awards that recognize the most extraordinary scientists will reflect existing trends in funding, because popular areas will provide a bigger pool of candidates from which to choose. But, she says, “I do agree that we should use the awards to drive change.”

EXPENSIVE GAMBLE

Although the Breakthrough Prize has been censured for playing it safe, critics are arguing that the Fundamental Physics Prize is making choices that are too risky. Woit notes that five out of the nine initial physics recipients — four of whom are based at the Institute of Advanced

NATURE.COM
Hear more about the
new prizes on the
Nature Podcast:
go.nature.com/eh6x3j

Study in Princeton, New Jersey — study string theory, the idea that elementary particles are composed of vibrating loops of energy. A vocal critic of string theory, Woit has long argued that this research area has attracted a disproportionate amount of funding, despite lying beyond the range of direct experimental testability. “The obvious danger is that you may be giving a large award to someone for a completely wrong idea,” he says.

Polyakov, whose own award was partly for contributing mathematical techniques to string theory, sees this willingness to gamble on ideas as the new prize’s niche. “For me, ideas have their own reality,” he says. But because in future the Fundamental Physics and Breakthrough prizes will be awarded by committees made up of the previous laureates, critics fear that current biases will be reinforced. Woit points to this year’s selection of Polyakov. “When string theorists at the Institute of Advanced Study give their first award to a colleague who works on the same stuff as them, then it is a serious problem,” he says.

Milner counters that next year the winners of the ‘special’ prizes — the seven Higgs hunters and physicist Stephen Hawking — will shift the balance on the judging panel. “It is a self-correcting loop,” he says.

PUTTING ON THE RITZ

The physics prize’s black-tie ceremony in Geneva, consciously modelled on the Oscars, highlighted the ambition of its founders to inspire current and future scientists. “I don’t see why millions shouldn’t ultimately watch this ceremony,” says Milner. Scientists in the audience were both entertained and bemused; one described it as “lots of fun”, another as “excruciatingly long”.

The question is, will the money and razzle-dazzle have any real impact? Stilgoe challenges Milner’s claim that the awards will encourage early career scientists to stay the course, rather than — like Milner, Zuckerberg and Brin — leaving for more lucrative pastures. “The idea that anyone would make a career choice based on the minuscule chance of winning, say, a Nobel, is ridiculous,” he says. “Scientists, on the whole, are not in it for the money — and I am not sure we should want them to be.”

Fred Cooper, a physicist and a visiting scholar at Harvard University in Cambridge, Massachusetts, also questions whether the awards will really speak to the public. “Visit YouTube and you’ll see that the public already turns to science celebrities — Michio Kaku, Brian Greene and Sean Carroll — to learn about physics, not to the winners of the Fundamental Physics Prize,” he says. “If outreach is your aim, then give money to those that are already great communicators.”

Even if the awards do inspire young people, Stilgoe argues that they send out the wrong message. “The prizes reinforce the mythology

of science in which lone geniuses come up with brilliant ideas on their own,” he says.

And some say that the prize money would be better used to drive research directly. In 2011, for example, Venter joined forces with the X Prize Foundation and health-care firm Medco Health Solutions, based in Northampton, UK, to offer a US\$10-million prize to the first team to sequence accurately the genomes of 100 centenarians. “I’m always more excited by awards that push or drive innovation, rather than ones that just recognize past achievements,” Venter says.

Many researchers favour the idea of targeting awards at promising scientists early in their careers. “A small award at this stage is a fantastic idea,” says Cooper. At this point, scientists

“I find it offensive that people are trying to either borrow the prestige of the Nobel, or buy it.”

are in a vulnerable position, struggling to win grants and often supporting young families. “It will just free scientists up to do more research — it’s about getting the biggest bang for your buck,” Cooper says.

This month, billionaire investor Len Blavatnik launched an award with prizes of \$250,000 for young scientists. The scheme, which builds on a previous, regional version that has been running since 2007, will be administered by the New York Academy of Sciences. (*Nature Publishing Group*, Wilczek and Venter are among the advisers for the award.) Blavatnik says that he saw a gap in the prizes market when he attended a Nobel ceremony several years ago. “What struck me was that most of the laureates were quite old and they received the award

FOLLOW THE MONEY

A crop of new science prizes (pink) offers winnings greater than the Nobels.

Name | Year introduced

Breakthrough Prize in Life Sciences (2013)
Fundamental Physics Prize (2012)

US\$3,000,000

Tang Prize (2013)

\$1,675,000

Queen Elizabeth Prize for Engineering (2013)

\$1,500,000

Nobel prize (1901)

\$1,200,000

Blavatnik Award for Young Scientists (2013)

Lasker Award (1946)

\$250,000

Fields Medal (1936)

\$14,700

for something they did 30 or 40 years prior to that,” he recalls. “I thought, in terms of impact on the world, it would be good to award young people and create something that would allow them to thrive.” One of the first winners of the regional Blavatnik Awards, Ruslan Medzhitov, an immunobiologist at the Yale School of Medicine in New Haven, Connecticut, says that the honour enabled him to attract funding and more prizes, including a share of the \$1-million Shaw Prize in Life Science and Medicine.

As for the new breed of mega-prizes, even some of the critics acknowledge — with a laugh — that they would accept one if it were offered to them. And Hans Clevers, a molecular geneticist at the Hubrecht Institute in Utrecht in the Netherlands and one of the inaugural

Breakthrough Prize winners, says that the new Nobels could rival the prestige of the old ones in 30 years or so, if they can consistently identify high-calibre winners.

The organizers of the Nobel prizes, however, remain unruffled by the upstarts. “For us, the important issue is to continue the good history and track record that we have,” says Lars Heikensten, executive director of the Nobel prizes, based in Stockholm. “If we fail, it will be because we fail to maintain that level of respect, not because other prizes are acting as rivals to us. We’ve been in this business for 110 years and we plan to be in it forever.” ■ [SEE EDITORIAL P.138](#)

Zeeya Merali is a science writer based in London.



FLY, AND BRING ME DATA

Unmanned aerial vehicles are poised to take off as popular tools for scientific research.

BY EMMA MARRIS

The Tempest — wingspan 3.2 metres, cruising speed 75 knots — was designed to fly into severe storms. But during a test run in March for a new project, it is soaring through the bluest of skies. On the ground below, PhD student Maciej Stachura of the University of Colorado (UC), Boulder, is tapping on a tablet computer, transferring control to the aircraft's own computer after a manual take-off. Systems engineer James Mack keeps his hands loose around a controller in case a problem arises, while Neeti Wagle, another PhD student, scans the skies to make sure the Tempest does not collide with anything.

The plane's job today is to locate a beacon sending out a simulated distress signal. As it circles overhead, the Tempest's gas-powered engine makes the distinctive lawnmower-like noise that calls to mind the informal name often given to such aircraft: drones. Unmanned aerial vehicle (UAV) is more commonly used in scientific circles.

The UC Boulder team watches and listens as the 40 minutes or so of flight time tick by and the Tempest becomes a distant speck in the bright sky. Then a note of concern enters Stachura's voice. "It is not doing a great job. It should be getting closer to us at this point," he says. Finally, the drone turns and heads back towards the beacon. "Oh, there it goes," says Stachura, clearly relieved.

The use of drones in science has taken a similarly roundabout route. NASA first experimented with custom-built UAVs in high-altitude research during the 1970s, but unmanned planes have been slow to catch on. Drones with top-notch sensors were too expensive to tempt researchers and cheap versions could not offer much of value. During the past decade, however, lower prices and technical advances — from on-board navigation using the Global Positioning System (GPS) to miniaturization of autopilots — have lured many scientific groups to experiment with UAVs.

Already, they offer an efficient way to gather data and are making important advances in polar research, volcano studies and wildlife biology. "They are on their way to becoming this indispensable and

revolutionary technology," says Adam Watts, an ecologist at the University of Florida in Gainesville who has flown drones for years.

But technical and legal hurdles stand in the way of their wider use. Researchers are trying to improve the autonomy, manoeuvrability and endurance of UAVs. And regulations, particularly in the United States, place strict limits on where and how researchers can use the devices. If these rules loosen up — and there are signs that they may — flying science robots may start taking to the skies in much greater numbers.

LOFTY HEIGHTS

The drones used by military forces to hunt down enemies have attracted growing scrutiny in recent years, but some of them have also been used for science. NASA has conducted hurricane and climate studies with Northrop Grumman's Global Hawk, which can reach an altitude of nearly 20 kilometres — much higher than commercial planes fly. The agency got the drone for free from the Air Force, but interested scientists must be prepared to pay US\$20 million for such a craft — no sensors included.

Most researchers have to make do with much smaller and cheaper systems. A radio-controlled fixed-wing UAV such as the Tempest can be bought off the shelf for a few thousand dollars. And quadrotor helicopters can be purchased for just \$300. Slap on a few sensors, an autopilot and a cheap computer preloaded with algorithms, and researchers have an unmanned aerial system (UAS).

Despite the differences in equipment, military and civilian drone-research programmes have been closely linked, with advances flowing between the two sides. Many university UAV programmes are, in fact, part-funded by the military.

For now, most researchers working with drones are focused on improving the technology to make the devices more agile, more autonomous and better able to work in groups. Autonomy requires a

➔ NATURE.COM
To see robot
quadrotors play the
James Bond theme:
go.nature.com/mmhn94



suite of algorithms to interpret data from sensors, make decisions about where to fly, control the plane's path and classify objects captured by the UAV's cameras. And all of that computing has to happen in real time on tiny, light computers bouncing around in three-dimensional space.

One area of focus is vision-based navigation. Systems that rely on the GPS can achieve little better than 3-metre resolution at best — fine for open outdoor landscapes, but not good enough for urban areas or indoor settings. Drone developers would like to send their machines into earthquake-damaged buildings to look for survivors, which would mean avoiding errant beams, power lines and closed windows. To do this, an aircraft requires a complex system of cameras, gyroscopes and accelerometers to figure out where it is — and where the obstacles are.

A team led by Ashutosh Natraj, now at the University of Oxford, UK, has taught drones with fish-eye cameras how to 'find' themselves. The robots' algorithms divide up the circular visual field into sky and ground, identify a horizon line between the two and then derive the drone's altitude and orientation. For city flying, the team is writing algorithms that recognize and use the verticals and horizontals of buildings and streets as guides to navigate up and down, forwards and backwards. At night, the drone could project a laser pattern onto its surroundings to orient itself. Camera-based navigation is smart, says Natraj, because a single camera collects more quality information than a number of expensive, heavy sensors such as laser range-finders, obviating the need to integrate many sensors. As part of a three-year project to design UAVs that can deliver medical care after natural disasters, Natraj is developing systems to do all the image processing on-board the helicopter, rather than through a wireless connection to a separate computer.

The Oxford disaster-relief UAV project is hoping to develop multiple UAVs that talk to one another. Such research on swarms is a hot area, says Hyunuch Shim, director of the Center of Field Robotics for Innovation, Exploration and Defense at the Korea Advanced Institute of Science and Technology in Daejeon, South Korea. "If you go fast, go

An autonomous research drone in Australia sprays herbicide onto weeds.

alone. If you go long, go together," he says. Data collection and search-and-rescue missions are faster and more efficient with a team of drones to pool data and provide redundancy in case some machines fail. But the use of more vehicles also adds complexity. Drones working together have to be able to communicate with one another and make collective decisions.

Researchers are also focusing on increasing the endurance of UAVs, most of which are fuelled by gas engines and batteries. To keep the weight and costs down, researchers often use tiny drones with limited fuel capacity, which means short flights. Some groups are working on miniaturizing batteries, others on making the planes smart enough to take advantage of thermal updrafts and wind features, as birds and gliders do. Roland Siegwart, head of the Autonomous Systems Lab at the Swiss Federal Institute of Technology Zurich, has a team developing solar-powered planes that would never have to land. "I call them 'low-flying satellites,'" he jokes. They could actually work better than satellites for collecting data, because researchers could direct them. "You can have an up-to-date image of bush fires, move them over illegal logging operations or look for people lost in the ocean," Siegwart says.

Researchers are also focusing on increasing the endurance of UAVs, most of which are fuelled by gas engines and batteries. To keep the weight and costs down, researchers often use tiny drones with limited fuel capacity, which means short flights. Some groups are working on miniaturizing batteries, others on making the planes smart enough to take advantage of thermal updrafts and wind features, as birds and gliders do. Roland Siegwart, head of the Autonomous Systems Lab at the Swiss Federal Institute of Technology Zurich, has a team developing solar-powered planes that would never have to land. "I call them 'low-flying satellites,'" he jokes. They could actually work better than satellites for collecting data, because researchers could direct them. "You can have an up-to-date image of bush fires, move them over illegal logging operations or look for people lost in the ocean," Siegwart says.

VIDEO STARS

Teams working on UAVs tend to keep abreast of each other's work through videos posted online. The field's biggest YouTube 'star' is Vijay Kumar at the University of Pennsylvania in Philadelphia. Kumar's group controls quadrotor helicopters indoors with a modified Vicon system — the motion-capture system used in Hollywood and by the video-game industry. His videos show drones flying in tight formation, transporting two-by-fours, and even — in one video with more than 3 million views — performing the James Bond theme on multiple instruments. "The Internet has changed the rules," says Shim. And, Siegwart says, "It also spreads the information a little further, which helps attract good students."

With new talent helping to make drones smarter and cheaper, the regulations that control unmanned flight will be the biggest barrier to their expanded use in research. “This is still the major issue,” says Siegwart.

That is particularly true in the United States, where Federal Aviation Administration (FAA) rules make it laborious to get permission to fly drones outside (except for non-commercial hobbyists, for whom the rules are looser). “We need permission to go out into a field on campus and fly something that is six inches off the ground,” grouses Eric Frew, head of the Research and Engineering Center for Unmanned Vehicles at UC Boulder. “It is a one-size-fits-all approach.”

The FAA, based in Washington DC, requires that would-be drone operators apply for and receive one of two certificates for their research programme if they want to fly their UAVs outdoors. The applications request a lot of information, “so the FAA can determine if the operation can be conducted without hazard to other aircraft or people and property on the ground”, according to the agency’s communications office. This means that certifications are not granted for flights in cities and other crowded areas. These certifications are also limited to a 20-mile-square area (around 32 square kilometres), so when the UC Boulder team took Tempest storm-chasing across swathes of the country, they needed 59 separate permissions.

Once a certification is obtained, usually within 60 days, a group can fly its aircraft during daylight hours in the designated spot for a year or two as long as they file a NOTAM — a Notice to Airmen — in advance with the FAA every time they want to fly.

Each flight also requires a certified pilot. During the March test flight, that was Stachura, who spent most of the test staring into the plane’s controls on his tablet. The FAA also requires that an observer be on hand to watch for potential collisions and that someone be monitoring the radio from the local airport.

FLIGHT DELAYS

Eric Johnson, who studies UAVs at the University of Georgia in Athens, has looked at regulations around the world and says that “among NATO countries, the United States is about the worst”. But as long as there are no accidents, the consensus seems to be that the regulations will loosen. The FAA Modernization and Reform Act, which passed last year, calls for the US Department of Transportation to produce a plan by late 2015 for “the safe integration of civil unmanned aircraft systems into the national airspace system”.

By contrast, says Johnson, Australia and Canada allow the most types of operation, perhaps because both countries have a lot of airspace and smallish bureaucracies. Salah Sukkarieh, who studies robotics and intelligent systems at the University of Sydney in Australia says that the country’s liberal regulations are allowing the UAS field to grow there, despite its funding being a fraction of that available to US scientists.

Although most drone research has focused on improving the UAVs themselves, some scientists have been putting the devices to use. In March, NASA used a small electric military drone, the Dragon Eye, to sample and photograph the noxious gas plume spewing from Turrialba Volcano near San Jose in Costa Rica. The team compared the Dragon Eye’s measurements of sulphur dioxide to those made by the Terra satellite in an effort to calibrate the space-based readings. It would have been too risky to send a human pilot near the volcano, where there are strong

updrafts and ash that could choke a plane’s engines.

James Maslanik, a remote-sensing expert at UC Boulder, has been involved in a number of studies using drones to measure various qualities of sea ice in the polar regions since 2000. Here, too, UAVs can venture into regions too dangerous for a manned aircraft. In the Arctic, Maslanik says, “we are flying these things at 100 feet off the ice, wind is 80 knots, temperature of minus 40 °C.”

At the opposite end of Earth, researchers from UC Boulder have used

UAVs to measure jets of wind that scream down from the Antarctic plateau into Terra Nova Bay. Such measurements could help scientists to understand the dynamics of sea-ice formation around Antarctica, which creates dense salty water that sinks and helps to drive global ocean currents. “Nobody had an aircraft out there during winter when the winds are strongest and took measurements because the conditions are too extreme,” says Maslanik. The data collected so far, he says, show unexpectedly complex wind patterns, including fierce, localized jets that push sea ice off shore and speed up its formation.

Biologists are also starting to use UAVs in their field work.

In India, the conservation group WWF is using drones to look for rhino poachers. Tom McKinnon, a retired engineer and managing director of InventWorks, a product-development firm in Boulder, is outfitting autonomous helicopters with nets to capture rare Mongolian vultures so that scientists can attach transmitters and study their movements.

On the plant side, Sukkarieh has developed a system using a fixed-wing UAV and a helicopter in tandem to locate weeds in remote rangelands and spray them with herbicide. And several groups are teaching drones how to tell one kind of plant from another, so that they can make maps of vegetation. Rather than purchase advanced sensors, which add weight and increase costs, Sukkarieh’s team is writing code to allow the UAV to map and classify vegetation using just GPS, a camera and an inertial measuring unit, which collects data on the position of the aircraft in space. The challenge of making trade-offs between sensors and weight has prompted Sukkarieh to think about designing UAV systems from scratch around their specific tasks, rather than just bolting sensors to an off-the-shelf aircraft. “What if the wings were sensors themselves?” he wonders.

For researchers without engineering expertise, however, the available UAVs offer plenty of opportunities. The Scottish Environmental Protection Agency, for example, purchased a drone in 2012 from Swiss company senseFly to survey estuaries for algal blooms — something that is difficult to do on foot. Susan Stevens, a scientist at the agency, says “you can get involved and use the technology without being an expert in it”.

Still, the best landings come with experience. As the UC Boulder team finishes up testing the Tempest, Mack, who was a UAV hobbyist before he joined this research team, gently sets the drone down on its belly in a patch of short grass. He picks it up in one hand to carry it back to the van.

Everyone is relaxed, having spent most of the 40-minute test flight doing little more than watching the Tempest and enjoying the spring day. If this is the future of field research, it looks pretty easy. “If everything goes well, it is fairly boring,” Stachura acknowledges. “Because it is autonomous, right?” ■

Emma Marris is a freelance writer in Klamath Falls, Oregon.



The Tempest in Colorado was designed to collect data from severe storms.

JACK ELSTON, UNIV. COLORADO

COMMENT



FACILITIES A call for new observatories to lead an Arab astronomy renaissance **p.161**

LITERATURE Robert Macfarlane reflects on the resurgence in nature writing **p.166**



BIOGRAPHY Graphic novel brings female primatology greats to teens **p.169**

CONSERVATION Drought is destroying Caatinga scrub forest in Brazil **p.170**

ILLUSTRATION BY PAUL BLOW



Stem-cell researchers must stay engaged

Recent developments have rekindled the ethical debate over human cloning. This is no time for complacency, caution **Martin Pera** and **Alan Trounson**.

Last month, news ricocheted around the world that reproductive biologist Shoukhrat Mitalipov and his colleagues at the Oregon Health and Science University in Portland had used cloned human embryos to produce stem-cell lines specific to individual patients¹. Although critics have since raised some problems with the paper, a preliminary enquiry indicates that the conclusions of the work still stand (see *Nature* <http://doi.org/mnk; 2013>).

This formidable technical feat is potentially a key step towards developing replacement tissues to treat disease. Media coverage of the paper has also rekindled long-standing controversies about human cloning, the use of human eggs and the destruction of human embryos. The achievement is a

timely reminder that scientists must remain actively engaged in discussions about the ethics of using human embryos for research in cell biology and regenerative medicine.

More than 1,000 embryonic stem (ES) cell lines have now been established worldwide. There has also been an exponential increase in the use of induced pluripotent stem (iPS) cells — a type of stem cell that can be made from reprogramming the cells of body tissues, such as the skin or blood, back to the embryonic state. As a result, many people question whether there is still a need to obtain ES cells from the ‘spare’ human embryos that are surplus to those needed for *in vitro* fertilization. In recent months, it seemed as if stem-cell research had finally moved on from the uncertainty over funding

and career prospects that has dogged it since 1998. Indeed, a prominent lobbying group for human ES-cell research, the Coalition for the Advancement of Medical Research in Washington DC, closed its doors last month (see go.nature.com/teedqq).

We believe, however, that enduring public concerns will inevitably resurface as stem-cell biology progresses. Also, new ethical challenges will need exploring — such as whether it is justifiable to produce human reproductive cells (or gametes) from iPS cells, and what they should be used for.

To ensure that rational discussion among scientists, policy-makers, regulators and the public precedes the formulation of regulatory policy, individual researchers should try to avoid confronting the public with ▶

► controversial scientific leaps out of the blue. Instead, scientists should gather to discuss the present and future course of human embryo research. They should also help to establish a formal programme for public consultation, similar to that led by two of the UK research-funding councils for synthetic biology. Ultimately, researchers should be prepared to continue arguing the case with governments and regulatory bodies when research on human embryos and ES cells is scientifically and medically merited.

FUTURE PROOF

Neither the current availability of numerous established ES-cell lines nor the increasing use of iPS cells rules out the possibility that human embryos will be needed in regenerative medicine in the future.

Clearly, the impetus for deriving new stem-cell lines from embryos has declined. More than 200 ES-cell lines are now on the US National Institutes of Health registry, meaning that anyone with the agency's funding can use them. Meanwhile, more than 1,200 lines are on the widely used International Stem Cell Registry, along with information about where to obtain them.

To produce new human ES-cell lines, researchers must negotiate a complex set of regulatory and compliance hurdles, obtain tens to hundreds of high-quality spare embryos, and do lab work that is labour-intensive, time-consuming and expensive. With so many well-characterized lines available, there is little incentive to derive new lines using established technology. In fact, an analysis in 2009 revealed that around 70% of the published research on human ES cells is based on two cell lines².

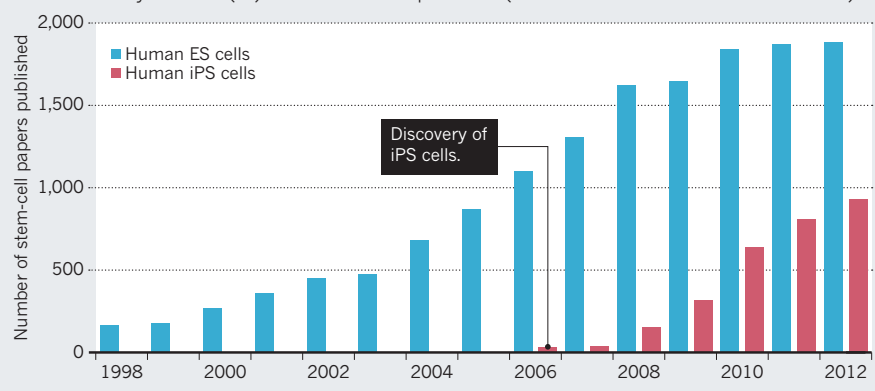
Yet the technology being used to derive and propagate cell lines does have shortcomings. For example, some cell lines are hard to renew and expand; about one-quarter of them develop genetic abnormalities *in vitro* after a routine period of cultivation³; and, in many cases, it is still very difficult to convert stem cells into fully mature functional cells, such as heart or liver cells.

Within the next decade, it is possible that technological improvements in the derivation and maintenance of ES cells will enable researchers to overcome deficiencies in culture systems. If this occurs, the use of new embryos to derive superior stem-cell lines might well be justified.

Similarly, it is unclear whether iPS cells will render ES cells redundant, despite the fact that the use of iPS cells has shot up in recent years (see 'Shifting preferences'). For example, there are still question marks over the genetic integrity of iPS cells and whether they differentiate into cells that are useful for therapeutic purposes as robustly as ES cells do. For instance, the skin cells of an adult, from which iPS cells can be derived, may have already

SHIFTING PREFERENCES

In recent years, research on human induced pluripotent stem (iPS) cells has grown rapidly, whereas studies of human embryonic stem (ES) cells seem to have plateaued. (Data include reviews and research articles.)



accumulated troublesome mutations, and the effects of reprogramming on the genetics of somatic cells are still being debated⁴.

Research groups in Japan and California are rapidly pushing studies on iPS-cell derivatives towards the clinic to treat age-related macular degeneration (a major cause of blindness) and genetic skin disorders, among other diseases. But the therapeutic potential of iPS cells relative to that of ES cells will be known only when the safety and efficacy of both cell types has been thoroughly evaluated in preclinical animal models and in early-stage clinical trials.

Even if iPS cells do make ES cells obsolete, new directions in research using iPS cells could fuel more lively debate among the scientific community, regulators and the public than has been spurred so far by work on ES cells. Healthy mice, for example, have been produced from fertilized eggs derived from iPS cells⁵. If it becomes possible to make human gametes from iPS cells, these could have many uses: to study the basis of human infertility; to identify factors present in the egg that might enhance its reprogramming for stem-cell lines; to produce embryos *in vivo* and *in vitro* for treating human sterility; or even to genetically modify the germline to prevent disease.

These possibilities may be even more ethically challenging than the use of spare embryos from *in vitro* fertilization to make ES cells. So far, such prospects have hardly been mentioned in the public arena.

CLONING REBORN

One procedure that has generated much ethical controversy is somatic cell nuclear transfer (SCNT), or 'therapeutic cloning'. Used by Mitalipov and his colleagues, this involves transferring the nuclear genome of an adult body cell, such as a skin or liver cell, into an unfertilized egg from which the nucleus has been removed. After 'tricking' the egg into becoming an embryo — by mimicking the chemical changes triggered by fertilization

— researchers can produce ES-cell lines that are genetically matched to the original donor.

Biologists initially envisioned SCNT as a way to produce patient-specific tissues needed for transplantations. Currently, it is permitted in only a handful of jurisdictions, including Britain, Australia, China, California, New York and Oregon. But until Mitalipov's breakthrough last month, no one had managed to convincingly produce ES cells from human cells using SCNT⁶.

The procedure is technically challenging, and a major stumbling block has been the need for numerous mature human eggs. In fact, interest in SCNT in humans waned substantially after the discovery of iPS cells, as measured by a decline in the number of SCNT papers and researchers working on it.

The findings of Mitalipov and his colleagues — or the future discovery of a way to derive hundreds of mature eggs from ES cells or iPS cells — could revive work on SCNT. Several biologists have argued that transferring nuclei from the somatic cells of humans to the eggs of another species, such as those from the frog *Xenopus laevis*, might be a powerful tool for understanding the reprogramming process in human cells⁷.

Even now, there is a compelling case for using SCNT-related technology in at least one clinical setting. Some mutations in mitochondrial DNA are associated with several potentially fatal disorders of the cardiovascular and nervous systems. Two studies from the past year show that the transfer of the haploid genome from an affected person into a healthy donor egg can prevent the inheritance of such mutations in cultured human ES-cell lines^{8,9}. And Mitalipov and his co-workers point out that SCNT-derived cell products could be used to treat patients with mitochondrial diseases. Such proof-of-concept studies may represent the best hope yet for patients, or for women who are at risk of passing on a mitochondrial disease to their children.

We think that SCNT should be permitted to facilitate experiments *in vitro*, but

not to develop a live human or human-animal hybrid. The use of SCNT-related technology to treat mitochondrial diseases does not involve cloning, but it does raise the question of whether it is acceptable for children to have three genetic 'parents': the mother who donates the egg nucleus, the father who donates the sperm nucleus and another woman who donates the mitochondrial DNA. The UK Human Fertilisation and Embryology Authority recommended in March that the government authorize the use of this technique to help these patients. We agree with this judgement.

Such possibilities need careful consideration and public consultation. We believe that the scientific community, which was forced to engage in ethical discussions in the early stages of stem-cell biology, should lead the way. As a first step, scientific academies such as the US National Academy of Sciences or the Australian Academy of Science should organize symposia to foster debate on the ethical ramifications of recent advances and possible new breakthroughs. Scientists should also engage with the public and the broader medical community; for instance, by collaborating with patient advocate groups such as the UK Juvenile Diabetes Research Foundation, and health-care providers such as the UK National Health Service. This would enable scientists to keep abreast of people's concerns, and to inform stakeholders of the realistic benefits and limits of their research and the ethical challenges it may bring.

The potential benefits of stem-cell research are immense. Prospects for transformative treatments for conditions such as macular degeneration, type 1 diabetes or Parkinson's disease are now on the horizon. But without first convincing governments, the public, and funding and regulatory bodies that all the possibilities have been thought through and evaluated, headline-catching results could create a backlash that unnecessarily delays the tremendous potential benefits of cell therapies. ■ **SEE NEWS & VIEWS P.174**

Martin Pera is at the University of Melbourne, Australia. **Alan Trounson** is at the California Institute for Regenerative Medicine, San Francisco, California, USA. e-mails: mpera@unimelb.edu.au; atrounson@cirm.ca.gov

1. Tachibana, M. *et al.* *Cell* **153**, 1228–1238 (2013).
2. Scott, C. T., McCormick, J. B. & Owen-Smith, J. *Nature Biotechnol.* **27**, 696–697 (2009).
3. The International Stem Cell Initiative *Nature Biotechnol.* **29**, 1132–1144 (2011).
4. Ronen, D. & Benvenisty, N. *Curr. Opin. Genet. Dev.* **22**, 444–449 (2012).
5. Hayashi, K. *et al.* *Science* **338**, 971–975 (2012).
6. Grieshammer, U., Shepard, K. A., Nigh, A. & Trounson, A. O. *Nature Biotechnol.* **29**, 701–705 (2011).
7. Narbonne, P., Miyamoto, K. & Gurdon, J. B. *Curr. Opin. Genet. Dev.* **22**, 450–458 (2012).
8. Tachibana, M. *et al.* *Nature* **493**, 627–631 (2013).
9. Paull, D. *et al.* *Nature* **493**, 632–637 (2013).



Observers in Amman, Jordan, watch the transit of Venus across the Sun in June 2012.

Time for an Arab astronomy renaissance

Arab Muslim countries need a new generation of observatories to rejoin the forefront of the field, says **Nidhal Guessoum**.

Islamic astronomy enjoyed a golden age from the ninth to the sixteenth century AD. Great observatories in Baghdad, Damascus, Maragheh, Samargand and Istanbul mapped the sky to set dates for religious and civil festivals and for astrology. Sophisticated calculations and models led to advances in mathematics.

Today, Arab astronomy barely registers on the world map. Scientific research is weak across the Arab world, and astronomy weaker still. Unlike countries of comparable gross domestic product per capita, such as Turkey, Israel and South Africa, most Arab nations are generating

fewer than ten papers in the field each year, and these are hardly cited. Few sizeable telescopes are operational or planned.

The lagging state of astronomy is a paradox for a region where funding should not be a serious constraint, at least in the wealthier Gulf states. The region has several excellent observing locations above 2,000 metres that benefit from clear skies. Public fascination is strong, as shown by the many local amateur associations and large gatherings for astronomical events, such as eclipses, comet passages or the most recent transit of Venus across the Sun in June 2012. ▶

► In my view, astronomy research is being neglected because of the strongly utilitarian Arab Muslim approach to science¹. Cultural principles, such as serving the people first, led Arab nations to build bases in the applied sciences in the second half of the twentieth century, including petrochemical engineering and pharmaceuticals. There was also a need for the region to develop its infrastructure quickly after the departure of colonial powers. Today, subjects such as theoretical physics are taught widely but are low cost and are considered low priority. Astronomy seems to require expensive buildings, equipment and technicians for little tangible return.

“Funding should not be a serious constraint, at least in the wealthier Gulf states.”

Another problem is the lack of expertise in the management of large scientific projects — an essential element if observatories and research centres are to operate effectively. The few large telescopes that have been built in the region in the past 50 years have been poorly run, are often inoperable and have produced few results.

I call on Arab countries to build a new generation of observatories. A few medium-sized telescopes (one- to two-metres in mirror diameter) costing a few tens of millions of dollars would allow Arab astronomers to join front-line research by searching for supernovae, the afterglows of γ -ray bursts, variable stars and extrasolar planets. Universities need to set up degree and international exchange programmes in astronomy to train and integrate the next generation of Arab astronomers. Such developments would galvanize academic and public interest in fundamental science across the region.

A GOLDEN PAST

Astronomy had a central place in society from the early times of Islamic civilization. In the early ninth century, a few decades after the founding of Baghdad as the capital of the new Muslim empire, the caliph al-Ma'mun (AD 786–833) ordered the erection of two observatories: Shammasiyya near Baghdad, and Jabal Qasiyun on the high outskirts of Damascus. Their main aim was to check solar and lunar data in old Greek and Indian tables, and to produce civil and religious calendars. Facilities included a quadrant made of marble with a radius of five metres to measure angles on the sky, and a sundial with a central gnomon — the column that casts the shadow — more than five metres high.

Islamic practice relies on astronomy for three purposes: computing prayer times for various locations and dates, which are based on the apparent motion of the Sun;

determining the direction to Mecca (the *Qibla*) for prayers; and establishing the dates for holy festivals, particularly Ramadan (the month of fasting) and Hajj (the pilgrimage), which are set by the observation of the thin crescent of the new Moon. All three still cause heated arguments among Muslim astronomers and scholars.

Historically, astronomy was also needed for navigation at sea and on land. Travelers and sailors learned that the arc of the Moon indicates the east–west line; the shortest shadow of a stick gives the north–south direction; the height of Polaris (the Pole Star) above the horizon gives the latitude of the place; and Mintaka, a star in Orion's belt, traces the celestial equator.

Muslim rulers were also guided by astrology, believing that some days were more propitious than others for mundane activities or stately decisions. Astronomers' ability to predict planetary motions and alignments, eclipses and new and full Moons was a powerful weapon in a ruler's arsenal. Courts had a resident astronomer, and mosques had a time-keeper (*muwaqqit*).

By the thirteenth century, rulers were erecting great observatories such as Maragheh (in present-day Iran), which was the largest in the world at the time. Astronomers and students from around the world used its sophisticated instruments, which included an armillary sphere model of celestial

body motions several metres wide, as well as its library of 400,000 books. Theories developed there include the 'Tusi couple' that links linear and circular motion, which was developed by the astronomer Nasir al-Din al-Tusi in 1247, and later used by Nicolaus Copernicus in his geometry of planetary orbits.

In the fifteenth and sixteenth centuries, even more stunning observatories were built. In the Samarqand observatory (completed in 1429; now known as Ulugh Beg Observatory), a 30-metre-high building housed ten instruments. These included an armillary sphere; an azimuthal quadrant for measuring the horizontal angle of the star from the north; and a meridian arc with a 40-metre radius, which measured celestial positions to within a few arcseconds. The Istanbul observatory, built in 1577, although smaller, also housed ten instruments and had 15 full-time astronomers². Sophisticated tables giving the positions of stars, planets, the Sun and the Moon were produced in each.

Thus hundreds of stars and constellations have Arabic names, such as Altair, Deneb, Vega and Rigel. Today, more than 20 lunar craters bear the names of Muslim astronomers, including Alfraganus (al-Farghani), Albategnius (al-Battani) and Azophi (al-Sufi). The scholar Abu Rayhan al-Biruni (AD 973–1048) used astronomy and trigonometry to determine Earth's circumference to within 0.3% of today's accepted value. Muslim women participated too: in the tenth and eleventh centuries, Fatima of Madrid, daughter of the great Andalusian astronomer Maslama al-Majriti, helped her father to produce tables of star and planet positions. In the tenth century, Mariam of Syria was a skilled constructor of astrolabes for celestial surveying.

From the thirteenth century onwards, major centres of learning were lost, such as those in the Iberian territory of Al-Andalus, and conservative rulers and clergy accorded religious knowledge an ever higher place than worldly science. Universities disappeared and old places of learning became antiquated and disconnected from scientific developments in Europe. Observatories were seldom gifted rich, religious endowments (*awqaf*) and thus rarely continued for more than a few years or decades after their establishment.

Thus the era of great Islamic observatories came to an end in the later part of the sixteenth century, with the demise of the Ottoman empire and the rise of European science. The practice of astronomy, as with other areas of science at the time, depended on the good will of the caliph or patron. The Istanbul observatory was destroyed in 1580, less than three years after its construction, by a new ruler who had been convinced by the religious



An eleventh-century astrolabe, used to measure celestial positions, among other functions.

BRIDGEMAN ART LIBRARY



The Ulugh Beg Observatory in Samarkand, Uzbekistan, completed in the fifteenth century, was used by several famous Islamic astronomers.

establishment that “prying into the secrets of the heavens” was reprehensible and would trigger God’s anger².

As a result, no astronomy and little science were conducted in Muslim countries until the late-nineteenth century.

ARAB ASTRONOMY TODAY

Things got going again when European powers — Britain and France, in particular — colonized many parts of the Arab Muslim world, bringing modern ideas with them, but education to only a select few.

For instance, the Lee AstroPhysical Observatory in Lebanon, named after its British merchant patron, Henry Lee, was built in 1873 by Cornelius Van Alen Van Dyck, a passionate professor of astronomy at what later became the American University of Beirut. The observatory housed a 25-centimetre telescope, which worked well enough until the facility closed in 1980.

In 1891, French astronomers built an observatory on the hilltops overlooking Algiers; it contributed 1,260 photographic plates of the sky between 1891 and 1911 to the Astrographic Catalogue project, a large international effort to map star positions to a high degree of accuracy. In Egypt, the Helwan Observatory was built in the early twentieth century^{3,4}; an astronomy

department was established at Cairo University, and the country joined the International Astronomical Union (IAU) in 1925 (ref. 5).

Sadly, a world map of today’s observatories shows just two medium-sized telescopes in Arab countries: Egypt and Algeria. By comparison, South Africa has half a dozen big observatories, including the South African Large Telescope (SALT) with an 11-metre primary mirror — the largest single optical telescope in the Southern Hemisphere. India has at least a dozen observatories, including the Indian Astronomical Observatory at Hanle, which houses a two-metre telescope.

The largest telescope to have graced the Arab world is the 1.88-metre instrument at Egypt’s Kottamia Observatory, in the desert 75 kilometres outside Cairo. The telescope was inaugurated in May 1964, but for decades it was under-used or broken. Refurbished in the 1990s, Egyptian astronomers say that the telescope is now working, although few papers have resulted from it.

In Iraq, an ambitious plan to build a world-class observatory in the northern high mountains was launched in the 1980s, envisaging 3.5- and 1.25-metre telescopes, along with a 30-metre radio telescope⁶. Wars and their resulting damage meant that the project was never finished. Plans to relaunch it have been aired, without progress.

In the past few years, two small observatories have been constructed in other parts of the Arab world. At an altitude of 2,750 metres, the Oukaimeden Observatory near Marrakesh in Morocco hosts a 50-cm robotic telescope for asteroid and comet searches. It is run by the Cadi Ayyad University in Marrakesh in collaboration with Uranoscope de l’Île de France (a French amateur astronomy association) and the Marrakesh Amateur Astronomy Association. Another observatory in Lebanon, built by Notre Dame University in Louaize, contains a 60-cm telescope, which is expected to begin operating soon. Other Arab countries have smaller telescopes, with mirrors of 35–50 cm.

Several Arab states have proposed one- to two-metre telescopes over the years, including Algeria, Libya, Oman, Saudi Arabia and the United Arab Emirates, but little progress has been seen.

RESEARCH ANALYSIS

To assess how badly astronomy research is suffering in the region, I compared publication and citation data for Arab nations with data from Iran, Israel, South Africa and Turkey (see ‘Arab astronomy papers’). Arab astronomers published fewer papers and had fewer citations than astronomers in those

PUBLICATION DATA

Arab astronomy papers

To assess the state of Arab astronomy research, I used the Thomson Reuters Web of Science to extract publication data for astronomy and astrophysics papers for each Arab country from 1 January 2000 to 31 December 2009. For comparison, I collected similar data for authors from Turkey, Iran, Israel and South Africa.

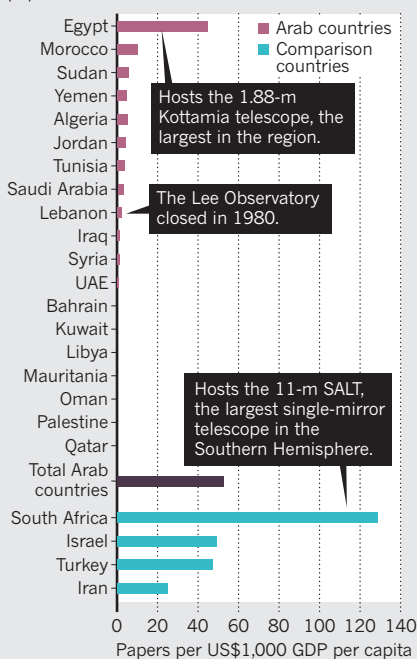
Because there were few papers for Arab countries, I examined them by hand and discarded ones on tangential and highly theoretical topics. The comparison countries had a greater number of papers, so I examined a random sample of 200 papers from each country and scaled the totals accordingly. For Arab countries, 40–50% of papers were excluded (reflecting the emphasis on theoretical work); for Iran, the percentage was 78%; for Israel, 25%; and for South Africa, 19%.

The number of astronomy papers as a proportion of science papers for the Arab world is 3 per 1,000 (ranging from 1 for Qatar to 6 for Bahrain; Yemen has an abnormally high ratio owing to its very low science production). This is similar to Iran (2) and Turkey (3), but much lower than Israel (14) and South Africa (24), the proportions of which are similar to those of the United States, China, India, Japan, Brazil and Spain. (In these countries, the range is 10–25 astronomy papers per 1,000 publications.)

The United Arab Emirates, for example, with a population of 8 million and a gross domestic product (GDP) per capita of US\$46,000 in 2011, published 6,000 science publications over 10 years, but only 23 of those were in astronomy. Israel, with a similar population but a 30% lower GDP per capita than the United Arab Emirates, published a total of 143,000 scientific papers, of which 1,500 were astronomy articles. Similarly, of the 13,000 science

PAUCITY OF PAPERS

Arab countries produce fewer astronomy papers than nations with similar GDPs.



papers that were published by authors from Lebanon, which has a population of 4 million and a GDP per capita of \$9,000, just 19 were in astronomy.

The citation figures are even more striking. For publications in 2000–09, there were 1,507 citations for papers that had a first author from an Arab country and 1,596 for papers that include an author from an Arab country (but not a first author). This is a total of 3,103 citations, compared to 4,355 for Turkey, which contains one-fifth of the population of the Arab world. Israel's and South Africa's citation figures were 20 times and 9.5 times higher, respectively, than those of the Arab world.

other four countries. The entire Arab world published fewer astronomy papers than Turkey alone, and substantially fewer than South Africa or Israel. Citation figures are worse: Arab astronomy papers were cited less often than Turkey's, South Africa's or Israel's.

As for degree programmes in astronomy or astrophysics at Arab universities, these can be counted on two hands. Small programmes exist in Egypt, Jordan, Lebanon, Saudi Arabia, Sudan and Algeria. Only a few dozen out of several million students major in astronomy or astrophysics at undergraduate or at master's level, and home-grown

PhD students are rare.

Conferences, colloquia and summer schools in astronomy are organized, but with modest academic impact. The Arab Union for Astronomy and Space Science (AUASS), a supranational organization linking professional astronomers and amateur associations of the Arab world, holds meetings every two years, but it has not published any proceedings.

LOOKING FORWARD

It is time for governments, funding agencies, science-advocacy organizations and universities of the Arab world to move beyond

the utilitarian view of science and promote professional astronomy.

Large projects in this field can inspire the science and technology community, the education sector and the public, and shift attitudes towards basic research in general.

This can be done by accelerating efforts to build high-class observatories, with one- and two-metre telescopes in several countries; establishing astronomy programmes in all public universities; setting up exchange agreements with international institutions; and funding graduate students to pursue doctoral programmes at universities abroad.

The Arab world offers ample sites for high-quality observatories — several mountains have peaks higher than 3,000 metres. Mountain ranges in the Arabian peninsula that span the United Arab Emirates, Oman, Saudi Arabia and Yemen typically enjoy 200–250 clear nights a year. Peaks in the Sinai peninsula reach up to 2,600 metres, where at least 150 summer nights are clear. Similar suitable sites exist in several other countries, from Iraq to Morocco.

Rich Gulf states could work together to set up a world-class observatory. A facility would cost between US\$50 million and \$100 million, including equipment (a two-metre telescope, photometer, spectrometer, fast computers and network links, and a weather station), buildings with work and meeting rooms, sleeping quarters and leisure areas, and local roads and infrastructure.

Arab universities should cooperate. Expert meetings should be convened to produce white papers on restarting astronomy in the region. These activities should be supported by international organizations such as the AUASS and the IAU to put pressure on governments. And it is essential that major Arab universities offer degree programmes in astrophysics.

Astronomy has a natural place high in the landscape of Arab Islamic culture. It must be brought back. ■

Nidhal Guessoum is professor of physics and astronomy at the American University of Sharjah, United Arab Emirates.
e-mail: nguessoum@aus.edu

1. Guessoum, N. *Nature Middle East* <http://dx.doi.org/10/mrn> (2012).
2. Sayili, S. *The Observatory in Islam* (Arno Press, 1981).
3. Hady, A. A. *Adv. Space Res.* **42**, 1800–1805 (2008).
4. Hassan, S. M. 'Kottamia Telescope Upgrading' in *Developing Basic Space Science World-Wide* (eds Wamsteker, W., Albrecht, R. & Haubold, H. J.) 237–240 (Springer, 2004).
5. Aiad, A. *IAU Colloq.* **105**, 398–399 (1990).
6. Al-Naimiy, H. M. K. *Proc. IAU Symp.* **260**, 429–437 (2011).



BRIDGET BESAW/CORBIS

ENVIRONMENT

New words on the wild

Robert Macfarlane reflects on the recent resurgence in nature writing.

“I don’t know anything else worth saving,” rages the hero of *The Monkey Wrench Gang* (1975) about the wilderness. Edward Abbey’s novel went on to inspire the 1980 formation of eco-activist group Earth First!, whose members undertook direct-action campaigns aimed at preserving wild places. “No compromise in defence of Mother Earth!” was the group’s original pledge.

Abbey’s novel was fuelled, like his memoir *Desert Solitaire* (1968), by fury at the despoliation of the landscapes he loved. The book formed part of an extraordinary surge of writing about nature in North America in the 1970s and 80s. Annie Dillard’s visionary *Pilgrim at Tinker Creek* (1974), about life and death in the Blue Ridge Mountains of

Virginia, won the Pulitzer Prize in 1975; 11 years later, Barry Lopez’s exploration of the American far north, *Arctic Dreams* (1986), became an instant classic and long-term best-seller.

Over the past 15 years, Britain has experienced a comparable surge with the blossoming of a literary form that has become known as ‘new nature writing’. The tone of this form, however, feels far from the roustabout activism of Abbey, or even the puckishness of Dillard. The genre is distinguished by its mix of memoir and lyricism, and specializes in delicacy of thought and precision of observation. A number of these books have drawn massive acclaim from critics and readers: Roger Deakin’s *Waterlog* (1999), a sparkling account of his ‘swimmer’s

journey’ through the lakes and rivers of the British Isles; William Fiennes’s *The Snow Geese* (2002), about flight, home and happiness; *Nature Cure* (2005) by Richard Mabey, who describes the role of the outdoors in his recovery from depression; and Olivia Laing’s *To The River* (2011), a fluent meditation on place and memory inspired by the River Ouse in Sussex.

It is widely agreed that Britain is going through a golden age of nature writing, but no one seems sure quite how to define it. Ragtag, wayward and polymorphous, it folds in aspects of memoir, travel, ecology, botany, zoology, topography, geology, folklore, literary criticism, psychogeography, anthropology, conservation and even fiction. Most distinctive, to my mind, is its tonal mix

of the poetic and the scientific and analytical. *The Snow Geese*, for instance, combines exquisite accounts of geese in flight — “the flock lifted from the field as a single entity, 10,000 pairs of wings drumming the air, as if people were swatting the dust from rugs” — with an inquiry into the biomechanics of avian migration.

Even this aspect of new nature writing, though, is hardly new. In the decade after the Second World War, US scientists Rachel Carson and Loren Eiseley, and pioneering conservationist Aldo Leopold — author of *A Sand County Almanac* (1949) — became famous for the intimate tone and ethical commitment of their essays. They were “imaginative naturalists”, to borrow the subtitle of Eiseley’s million-selling *The Immense Journey* (1957). Their appeal lay in their fusion of the latest research with first-person narrative.

CONTINENTAL DIVIDE

North America has always been a happier habitat for nature writing than Britain, perhaps because the vast and various geography of the continent — from southern canyon lands to northern polar tundra — has provided limitless inspiration. For much of the twentieth century in the UK, writing about wildlife or the countryside was regarded with suspicion tending to contempt. Stella Gibbons parodied rural writing in *Cold Comfort Farm* (1932): “Daisies opened in sly lust to the sun-rays and rain-spears, and eft-flies, locked in a blind embrace, spun radiantly through the glutinous light...”. And Evelyn Waugh skewered the plush prose of country diarists in his novel *Scoop* (1938), with sentences such as “Feather-footed through the plashy fen passes the questing vole”.

For decades after Gibbons and Waugh, nature writing seemed in abeyance, dismissed as either reactionary or soppy in its longing for oneness with the natural world. This didn’t prevent the emergence of occasional masterpieces: J. A. Baker’s *The Peregrine* (1967), for instance, inspired — as in part was Carson’s *Silent Spring* (1962) — by Derek Ratcliffe’s landmark studies into the effects of pesticides on eggshell-thinning in British raptors. There was also Nan Shepherd’s glorious meditation on place and being, *The Living Mountain* (1977), born of years of acquaintance with the Cairngorm massif of north-east Scotland.

Why, then, has nature writing enjoyed its recent renaissance in Britain? Two main causes suggest themselves. The first is disembodiment: people are spending increasing amounts of time in atmosphere-controlled environments, hunched at keyboards. An inevitable consequence of this has been a longing for wildness and nature: the feel of wind and sun upon the face,

or the sight of a stooping falcon, or of an oak tree in spring leaf. Or, as Henry David Thoreau cried — having laid his hands on the summit rocks of Mount Katahdin in Maine — for “Contact! Contact!”

The second context is global crisis. It is no coincidence that a literature celebrating the natural world should have emerged at a time when the natural world is so conspicuously under threat. The past 15 years have seen the Deepwater Horizon blowout in the Gulf of Mexico, the break-up of the Antarctic ice shelf, widespread habitat destruction, further evidence that we are living through the sixth great extinction pulse, and the slow-motion emergency of climate change.

British nature writing is energized by this sense of menace and hazard. The biodiversity crash has been tackled boldly by Melanie Challenger in *On Extinction* (2011), and wittily by Caspar Henderson in *The Book of Barely Imagined Beings* (2012). Henderson records some of the extraordinary species with which we share the planet (including the mantis shrimp *Gonodactylus*, which “has the fastest genitals in the West and will use them to smash your head with massive force”). Meanwhile, Esther Woolfson’s *Corvus* (2008) and *Field Notes From A Hidden City* (2013) examine interspecies relationships and the responsibilities we bear to the creatures that surround us (echoing Leopold’s ‘land ethic’).

A sense has emerged that nature writing might almost provide salvation. “If we are to continue to live with birds about us we need bird poems as much as the RSPB [Royal Society for the Protection of Birds],” notes ornithologist Tim Dee, author of *The Running Sky* (2009). And Margaret Atwood hopes her novels might “move public opinion in a more biosphere-friendly direction”. Nature writing to the rescue!

I am less sure. The Anthropocene is a frightening and forceful era. I wonder, really, what literature can do in the face of population pressure, rising sea levels, deforestation and the rapacious instinct of capital. A few years ago a monograph by the American academic John Felstiner, was entitled *Can Poetry Save The Earth?* (2009). “No!” I yelled silently at the cover when I first saw it, “Of course it can’t!” On reading the book, I was curiously relieved to find that this was also Felstiner’s answer.

There are good reasons not to exaggerate the possible consequences of nature writing. One is that it often preaches to the converted: the people it reaches tend to be those with the most developed environmental consciences. “Sceptical readers,” as the American farmer and writer Verlyn

Klinkenborg observed, “seldom pick up this kind of writing, or submit to its evidence.” Another reason is that nature writing can often feel too pious and gentle in its urgings — the green equivalent of attending Sunday school.

And yet a law of unintended consequences has always governed literature, as *The Monkey Wrench Gang* made clear. Writing that aims to provoke specific behaviours has a name: propaganda. And writing that seeks to provoke specific emotions also has a name: kitsch. Literature, by contrast, does not deal in deliverables. It stirs the sediments of thought and morals, setting them strangely awirl.

“Transformation comes about as much because of pervasive changes in the depths of the collective imagination as because of visible acts, though both are necessary,” says Rebecca Solnit, one of today’s most interesting US essayists and environmental activists. “And though huge causes sometimes have little effect, tiny ones occasionally have huge consequences.” The history of environmental literature is rich with such fascinating ‘transformations’: naturalist John Muir’s influence on the founding

of US national parks, or the thunderclap publication of *Silent Spring*, with its consequences for use of the pesticide DDT (dichlorodiphenyl-trichloroethane). As yet unmapped is the influence of Cormac McCarthy’s *The Road*

(2006), a novel that chilled me to the core and that British journalist George Monbiot has described as “the most important environmental book ever written”.

For literature possesses certain special abilities, very different to those of science. It can convey us into the minds of other people, and even — speculatively — the minds of other species. It can help us to imagine alternative futures and counter-factual pasts. It is content with partial knowledge in ways that science is not. Crucially it can, in author and environmentalist Bill McKibben’s phrase, make us feel things “in the gut” — fear, loss and damage, certainly, but also hope, beauty and wonder. And these last are, I think, the most important emotions in terms of our environmental future: our behaviour is more likely to be changed by promise than by menace. We will not save what we do not love. ■

Robert Macfarlane is a fellow of Emmanuel College, Cambridge, UK, and author of *Mountains of the Mind*, *The Wild Places* and most recently, *The Old Ways*.

► NATURE.COM

To hear Robert Macfarlane talk about travel writing: go.nature.com/bnuclea

TECHNOLOGY

Built by bicycle

Andrew Robinson mulls over a study of India's adaptation of low-tech inventions.

In an image taken by the great photographer Henri Cartier-Bresson, two Indian men walk down an empty, rural, palm-fringed road. Between them is an old-fashioned bicycle. One man grips the tip of a large, conical metal object perched on the saddle — the nose cone of a small rocket. The caption reads: "Near Trivandrum, Kerala. 1966. Preparing for a launch at the Thumba Rocket Equatorial Launching Station, housed in a former church."

That photograph of two space scientists encapsulates the thesis of *Everyday Technology*. This pioneering study by historian David Arnold examines India's response to certain small-scale technologies from the 1880s through to independence in 1947 and up to the 1960s — long before the country's digital revolution.

During the colonial era, British officials in India tended to regard its population, particularly in rural areas, as too mired in conservatism, poverty and illiteracy to adopt new technologies. So officials preferred to introduce large-scale technological projects from the top down, for example electric telegraphs, railways and irrigation schemes. After 1947, this attitude influenced India's first prime minister, Jawaharlal Nehru, who conducted an all-out pursuit of foreign-constructed hydroelectric dams and steel mills, and introduced nuclear power and a space programme. For Nehru, big dams were "temples of the new age" — emblems of an India untainted by its messy social reality.

Arnold, by contrast, believes that understanding technology demands an appreciation of the society embracing it, "even when the technological goods themselves remain largely foreign". He also argues that the slow spread of small-scale technologies, such as the sewing machine, prepared the way for India's later adoption of more sophisticated ones. By domesticating imported inventions, colonial societies undergo self-transformation, Arnold suggests.

As Cartier-Bresson's photograph hints, the bicycle in India has been a means of carrying people (sometimes three or four at a time), things and ideas. Even today it remains essential for millions of poorer Indians, who now may well also use a mobile phone. Bicycles have been converted into

cycle rickshaws and three- and four-wheeled carts; their basic mechanism has been used to power knife grinders and foot-powered looms. As Arnold shows, by the early twentieth century, the bicycle — along with three other low-tech mainstays, the sewing machine, the typewriter and the rice mill — were deeply woven into Indian society.

In the early 1920s, India was importing almost 50,000 bicycles a year; by independence, the number was five times that. In 1948, during the final fast of Mahatma Gandhi, the great leader of India's independence movement, 5,000 cyclists converged at a house in Delhi to hear Nehru report on Gandhi's health, their cycle lamps glowing



Henri Cartier-Bresson's Indian space scientists.

in the twilight garden like giant fireflies.

Most of these technologies were opposed by Gandhi. But, as Arnold is at pains to detail, Gandhi's well-known aversion to machines was not down to Ludditism. It was based on serious thought, and had a strong influence on the development of the post-1947 cottage industries movement.

Gandhi opposed the bicycle mainly because buying an imported luxury would lead to debt, although he permitted himself to travel by automobile. He had employed typists in his legal practice in turn-of-the-century South Africa and learned to type. But when he returned to India in 1915, he declared the typewriter "a cover for indifference and

Everyday Technology: Machines and the Making of India's Modernity

DAVID ARNOLD
Chicago University Press: 2013.
224 pp. £21

laziness", preferring to write his voluminous output of letters and articles by hand.

He objected to the rice mill — firstly because it would deprive poor women of income from pounding rice, secondly because pounding was good exercise and thirdly because milling removed the vitamin thiamine from the pericarp of the rice grain, a deficiency of which causes the disease beriberi that affected parts of India. Often critical of Western medicine, Gandhi was happy in this case to accept scientific evidence from two colonial nutritionists, Robert McCarrison and W. R. Aykroyd, but ignored their argument that less rigorous milling would preserve sufficient levels of the vitamin to prevent beriberi. However, Gandhi famously advocated the spinning wheel, and (less famously) championed the treadle sewing machine, particularly Singer's, describing it as "one of the few useful things ever invented".

Everyday Technology organizes an enormous amount of unfamiliar detail on a hitherto largely neglected subject, reinforced with copious statistics and illustrated with some appealing historical and contemporary images. It is enlivened by apt quotations from novels and films of the period, although regrettably includes none from the films of India's greatest director, Satyajit Ray. Ray's works offer many subtle reflections on people and technology, not least the trains, small-scale machinery and office atmosphere depicted in his celebrated *Apu Trilogy*.

However, the parts of this book are greater than the sum. The author's thesis is abundantly proven, but his conclusions seldom surprise. I am also left with the uncomfortable feeling that for all the enthusiasm with which modern India has responded to foreign technology, it has yet to create anything comparable with the achievements of its pre-colonial mathematicians, scientists and technologists. ■

Andrew Robinson is a writer and the editor of *Exceptional Creativity in Science and Technology*. He has written nine books on India, and the forthcoming *India: A Short History*.
e-mail: andrew.robinson33@virgin.net

HENRI CARTIER-BRESSON/MAGNUM

➔ NATURE.COM

For Andrew Robinson on India's digital boom, see: go.nature.com/ncddfa

Dian Fossey viewing mountain gorillas in *Primates*.

ETHOLOGY

Primatological derring-do

Kelly Stewart revels in a graphic biography that follows the human and scientific stories of three iconic primate researchers.

As enthralling careers go, those of trailblazing primatologists Jane Goodall, Biruté Galdikas and the late Dian Fossey take some beating. Now the personal and scientific stories of these pioneers of research on, respectively, wild chimpanzees, orangutans and gorillas feature in *Primates*, an engaging graphic biography by Jim Ottaviani and illustrator Maris Wicks. Unifying the three intimate first-person narratives is the figure of renowned palaeontologist Louis Leakey, who helped to launch all three researchers' careers.

Ottaviani clearly carried out extensive research on published material by and about the trio of researchers, including diaries and letters. For a book that takes well under an hour to read and aims to engage teens, *Primates* offers a remarkable amount of information on many different levels. The life stories may be rendered as cartoons, but the characters come across as multi-dimensional. And there is plenty of human interest, from Fossey's uncompromising ferocity and mercurial personality to Galdikas's painful choice between returning to Canada with her husband and young son or remaining in Indonesia to continue her work with orangutans.

How true to life is it? As someone familiar with all three stories — especially that of Fossey, with whom I studied gorillas in Rwanda — I'd say it's an accurate rendition.

We learn about the logistics of fieldwork, which involves no shortage of discomforts to satisfy the gruesome fascination of young readers — days of being rained on, isolation, exhaustion and illness. Galdikas's story, set in the leech-infested Indonesian forest, is especially rich in the 'ick factor'. The characters cut trails, sift through dung for food remnants, spend months living alongside and observing their subjects, present their findings at a conference, suffer academic insecurity and social awkwardness, and struggle to balance anthropomorphism and objectivity.

Making it all worthwhile is the fascinating allure of living in the wild and becoming immersed in the lives of members of another species. Triumph also comes with discoveries, such as Goodall making the first observation of chimpanzees using tools. And of course, Ottaviani describes the inevitable fight to conserve the apes and their habitats, which eventually becomes the crusade of all three scientists.

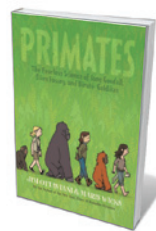
The narrative occasionally suffers from a

confusing shift in storyline or viewpoint, but is generally pithy and fast-paced. Although Wicks's artwork does not dazzle, it gets the point across, making ample use of expressions of comic exaggeration such as bug-eyed surprise. She also applies comic touches that convey the vitality of forests and their denizens, with ape vocalizations that change colour and burst out of the cartoon frame. Many illustrations are clearly modelled on specific photographs from early *National Geographic* articles.

The best thing about *Primates* is that Leakey and the primatologists essentially become action heroes. They are unconventional but undaunted, persevering against the odds. They make sacrifices, lead daring lives, uncover mysteries and fight for the good; science and scientists are portrayed as cool and exciting.

This book won't teach kids much about the great apes, but that isn't its point. If it inspires young readers to explore the reading list provided at the end, and perhaps become scientists or conservationists, then — as one might say at the end of any action comic — 'mission accomplished'. ■

Kelly Stewart is a research associate in the Department of Anthropology, University of California, Davis. She is co-author of *Gorilla Society* with Sandy Harcourt. e-mail: kjstewart30@gmail.com



Primates: The Fearless Science of Jane Goodall, Dian Fossey, and Biruté Galdikas

JIM OTTAVIANI AND
MARIS WICKS
First Second: 2013
144 pp. \$19.99

Correspondence

Save Caatinga from drought disaster

Brazil's semi-arid Caatinga scrub forest is experiencing its worst drought in 30 years, with more than 300 settlements in the northeast at the point of collapse (see go.nature.com/pngjfq). The federal government must urgently address the drought's disastrous effects on livelihoods and on the survival of this biosphere reserve.

Whether natural fluctuations in temperature and rainfall or climate change are to blame, the lack of water is killing livestock and destroying crops. Pressure on Caatinga land is increasing as local people hunt wild animals for food and trade, often burning vast areas of vegetation to flush out their prey. We have seen native plants being used indiscriminately to fuel furnaces for brick production so that families can buy food and water from other regions of Brazil.

The plight of this stricken area is being largely ignored by the media and by the government. Human survival should no longer need to depend on the destruction of local biodiversity.

Roberto Leonan Morim Novaes, Saulo Felix *Federal University of the State of Rio de Janeiro, Brazil.*
roberto_leonan@yahoo.com.br
Renan de França Souza *State University of Rio de Janeiro, Brazil.*

In praise of open research measures

On behalf of the Data-Enabled Life Sciences Alliance (DELSA Global), we applaud the significant, timely steps *Nature* is taking to ensure reproducibility and transparency in life-sciences articles (*Nature* 496, 398; 2013 and go.nature.com/oloeip).

We discussed *Nature's* Reporting Checklist for Life Sciences at our annual workshop last month (see www.delsaglobal.org). By encouraging researchers to make their data and metadata available, and to clarify their

analysis methods, the checklist will help to prevent mistakes from being propagated and resources from being wasted on dead-end experiments.

This is important in an era of tight funding and limited training in the quantitative aspects of research, both of which inhibit confirmatory experimentation. In addressing the veracity of data as well as the reliability and reproducibility of research, *Nature's* checklist will stimulate the transformation of data into knowledge, action and outcomes.

Scientific advances need strong public support to make a difference, and your policies constitute an important step in preserving public trust in science. The checklist can act as a useful template for development by publishers, federal agencies, funders, research organizations, societies and communities.

Eugene Kolker* *Seattle Children's Research Institute, Seattle, Washington, USA.*
eugene.kolker@seattlechildrens.org
*On behalf of 21 co-signatories. See go.nature.com/6mymnw for full list.



Curb China's rising food wastage

China is currently managing to feed its people (F. Zhang *et al.* *Nature* 497, 33–35; 2013), but food loss and waste throughout the supply chain must be taken into account if food security is to be maintained in the future.

Of China's grain output, an estimated 8%, 2.6% and 3% are lost during storage, processing and distribution, respectively — a total of some 35 million tonnes annually. As in many other developing countries, these alarming losses are a result of inadequate infrastructure, knowledge and technology, and are exacerbated by a decentralized agricultural production system.

China's increasing affluence is also leading to wide-scale food wastage. For example, household food waste totals roughly 2.5% of grain a year (around 5.5 million tonnes). This is fast approaching Western levels (see, for instance, go.nature.com/erz4if).

The pattern and scale of food waste in China are still unclear:

more quantitative research will help to inform policy-making and to increase public awareness of the problem (see our pilot study at go.nature.com/8zq1je; in Chinese).

Gang Liu *Norwegian University of Science and Technology, Trondheim, Norway.*
gang.liu@ntnu.no
Xiaojie Liu, Shengkui Cheng *Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.*

Education: enticing students into science

Colin Macilwain argues that the United States would not need to spend US\$3 billion annually on a programme to encourage young people to pursue careers in science, technology, engineering and mathematics (STEM) if market forces were right (*Nature* 497, 289; 2013). But conditions and opportunities in science have not visibly improved for students in the past three decades.

Twenty-four years ago, the US government was also trying to attract young people into STEM (T. Packard *Eos* **70**, 709; 1989). Then, 66% of the ocean-science community was living hand-to-mouth on short-term government grants. A university professor was expected to do 40 hours of teaching and administration a week, and 40 hours of research. Researchers who did not receive funding from their universities dared not spend time away from their work, lest their publication record should drop.

It was clear to me at the time that if research centres, universities, governments and societies wanted more people to work in science and technology, then salaries, job stability and job security would have to improve. They still have not.

Young people continue to shun research and instead opt to use their mathematical skills in accounting, their analytical skills in investment banking and their love of science in medicine. Macilwain blames business for the woeful range of scientific opportunities available to graduates. Whether the fault lies with business, government or universities, the educational pipeline in science and engineering does not work because graduates are scared off by what they see as a meat grinder at the other end.

Theodore T. Packard *University of Las Palmas de Gran Canaria, Spain.*
theodore.packard@ulpgc.es

Education: science literacy benefits all

Colin Macilwain wields too wide a brush in painting US federal funding of STEM education (for promoting 'science, technology, engineering and mathematics') as having the sole purpose of bolstering the workforce (*Nature* **497**, 289; 2013). This funding also achieves general science literacy, particularly when it is directed towards children in primary and secondary education or undergraduate students.

No matter how far they are pushed, most teens and young adults will not become scientists.

Fortunately, many STEM programmes familiarize students with the scientific process and with the natural world. Learning fundamental concepts also teaches them how to interpret and handle scientific information.

Science literacy subsequently benefits individuals throughout their lives, from forming opinions about proposed government policies to making health-care decisions. A well-informed citizenry, in turn, pays dividends to society as a whole.
Aaron C. Hartmann *University of California, San Diego, USA.*
achartma@ucsd.edu

Climate and war: a call for more research

The possibility that climate change could be responsible for violent conflict (A. Solow *Nature* **497**, 179–180; 2013) is starting to influence how governments frame and react to climate change. However, a real problem in this area is a paucity of theory to explain the associations (if any) between climate change and the outbreak of violence.

One overlooked factor is that populations caught up in conflicts or living in post-conflict societies are often more vulnerable to climate change. For example, the presence of landmines makes productive land inaccessible.

Climate policies can themselves be a source of conflict (see go.nature.com/zutmox). Measures that manage carbon sources and sinks or treat them as commodities — such as land-use changes, hydropower development or initiatives to reduce emissions from deforestation — can stimulate civil unrest if implemented without adequate checks.

Poverty, a history of fighting, and weak governance are well-established risk factors for conflict. The likelihood of violent conflict is reduced by democracy, social protection, effective justice systems and the protection of property rights. The influence of climate change on these factors warrants further investigation to guide policy-makers in promoting peace and prosperity in a changing climate.

Neil Adger *University of Exeter, UK.*
n.adger@exeter.ac.uk
Jon Barnett *University of Melbourne, Victoria, Australia.*
Geoff Dabelko *Ohio University, Athens, USA.*

Climate and war: no clear-cut schism

We are sceptical about the effectiveness of Andrew Solow's proposals for cooling the debate over a possible link between wars and climate change (*Nature* **497**, 179–180; 2013). We think that the division between the two sides ('quants' versus 'quals') is not as clear-cut as he implies.

Solow argues that this dividing line distinguishes between those who search for connections between violence and natural phenomena, including climate-related factors (quants), and those who prefer to explain conflicts as social processes (quals). But both approaches are studied by quants as well as quals. Quants may study climate-related effects on conflicts by analysing single events in detail or by considering many wars on aggregate using statistics.

There are also strong disagreements among those on each side of Solow's dividing line. For example, quants as well as quals include both proponents and sceptics of the connections between climate change and violent conflict.

In our view, the true divide is not so much about substance as about perspective.

Michael Brzoska, Jürgen Scheffran *University of Hamburg, Germany.*
brzoska@ifsh.de

Gender equality in Australian academies

Women are not under-represented across all learned academies in Australia (see *Nature* **497**, 7 and *Nature* **497**, 439; 2013). For example, the Australian Academy of Technological Sciences and Engineering (ATSE; of which I am president) has taken steps to ensure that women are appropriately recognized and included in all its activities.

Gender imbalance can adversely affect all stages of a scientific career, from tertiary education to employer recruitment, retention and promotion, with implications for a country's productivity and prosperity. Over the past three years, ATSE has led the way in identifying and promoting female talent across the science and technology sector in Australia, and within the academy itself.

One key element of ATSE's gender-equality policy is to identify women candidates for fellowship nomination through active search and mentoring processes. Last year, 10 of 37 elected fellows were female, and women now comprise 40% of ATSE's governing board.

Alan Finkel *ATSE, Toorak, Australia.*
alan@finkel.net

European concerns over GM salmon

As investigators for the European Food Safety Authority into the environmental risks posed by genetically modified (GM) fish, we are concerned about the US Food and Drug Administration's imminent approval of GM salmon (*Nature* **497**, 17–18; 2013).

This is a huge step that could encourage aquaculture of other GM fish in other countries, and not necessarily under strictly biosecure conditions.

There is still considerable uncertainty surrounding the environmental and physiological effects of escaped, fast-growing GM fish on aquatic systems. This reflects a poor understanding of how different species might be affected as the modified gene is expressed in the wild.

Europe's regulatory guidelines for aquaculture of GM fish and other alien species in Europe will therefore be underpinned by rigorous risk assessment (see go.nature.com/p6x2qb).

J. Robert Britton *Bournemouth University, Poole, UK.*
Rodolphe E. Gozlan *Bournemouth University, Poole, UK; and Institut de Recherche pour le Développement (UMR 207), Paris, France.*
rudv.gozlan@ird.fr

STEM CELLS

Cloning human embryos

Human embryonic stem cells have at last been generated by a technique called somatic-cell nuclear transfer. Further research on such cells should provide insight into ways of improving the generation of stem cells by reprogramming.

CHRISTINE L. MUMMERY &
BERNARD A. J. ROELEN

The birth of Dolly the sheep in 1996 produced great excitement among researchers. This first cloned mammal had been created¹ by introducing the nucleus of a somatic (non-germ) cell into an egg cell from which the genomic DNA had been removed, and transferring the resulting embryo into a foster mother. One implication of this achievement was that similarly cloned embryos could be used to produce stem cells that would be genetically identical to the cells of the somatic-cell donor, so that if these stem cells, or cells or tissues derived from them, were transplanted into the donor for treatment purposes they would not be rejected by the donor's immune system. In the years that followed, this technique of somatic-cell nuclear transfer (SCNT) was successfully used to produce stem cells from cloned mouse embryos^{2,3}, but all attempts in humans had failed — until the publication of a study in *Cell* by Tachibana *et al.*⁴.

The study is noteworthy for several reasons. First, it explains why all previous attempts at cloning human embryos have failed. The human egg (oocyte), and that of most mammals, is released from the ovary at the metaphase II stage of meiotic cell division. The cell resumes and completes meiosis only after it is fertilized. Removal of the meiotic spindle — the cellular structure that ensures the faithful distribution of chromosomes between dividing cells — is an integral part of SCNT. Tachibana *et al.* realized that this induces premature completion of meiosis in human eggs and subsequent loss of their capacity to reprogram somatic cells to a pluripotent state, which would allow them to differentiate into all cell types in the body. Crucially, the addition of caffeine to the culture medium slowed meiotic completion, ensuring the success of the authors' procedure.

The paper also shows that blastocysts (roughly 100-cell embryos) derived using this modified SCNT protocol were healthy enough to be used for generating embryonic stem (ES) cells that were genetically identical to the donor nucleus. And, notably, the authors have managed to generate these SCNT-ES cells using nuclei not only from fetal cells but also

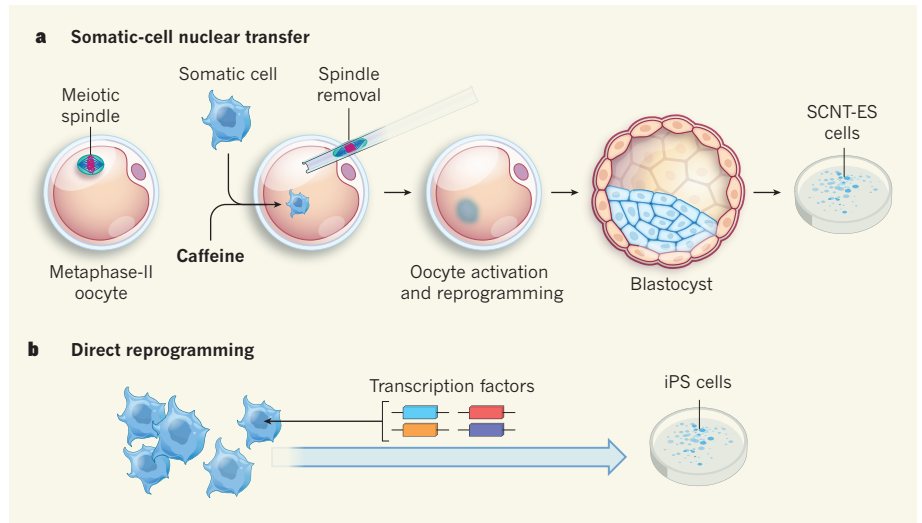


Figure 1 | Generation of pluripotent stem cells *in vitro*. **a**, Tachibana *et al.*⁴ show that human embryonic stem (ES) cells can be generated by a technique called somatic-cell nuclear transfer (SCNT). The authors removed the meiotic spindle from an oocyte arrested at the metaphase II stage of meiotic cell division. The oocyte had been incubated with caffeine to prevent premature completion of meiosis. They then inserted a somatic cell into the enucleated oocyte. Oocyte activation and cellular reprogramming followed, leading to blastocysts from which SCNT-ES cell lines were derived. **b**, By comparison, the generation of induced pluripotent stem (iPS) cells involves the introduction of four pluripotency-related transcription factors into a differentiated cell to induce its direct reprogramming.

from post-natal cells. This approach could therefore be used to create cellular models of the genetic disease that a somatic-cell donor might carry.

These technical achievements, for which researchers worldwide have strived for at least a decade, should be celebrated. Nonetheless, parallel advances made in the field of stem-cell research somewhat dampen the excitement that the present paper might have received — the sort of excitement that was generated some nine years ago by Woo Suk Hwang's report of similar results, before it was discovered that those data had been fabricated⁵.

In our opinion, the discovery⁶ in 2006 that differentiated adult cells can be directly reprogrammed to a stem-cell-like state called induced pluripotent stem (iPS) cells was a more significant breakthrough for this research field. iPS cells can be generated by introducing just four transcription factors into differentiated cells of an individual, without the need for the ethically sensitive step of creating embryos from oocytes as intermediates (Fig. 1). Indeed,

many laboratories now routinely generate iPS cells from patients, bypassing the practical and regulatory difficulties associated with obtaining human oocytes.

But an intriguing question now is how similar human iPS and SCNT-ES cells are. One difference is immediately apparent. In iPS cells, mitochondria (organelles that are the main source of cellular energy), as well as all other organelles, originate from the donor cell. In SCNT-ES cells, the mitochondria are derived from the oocyte and not from the donor of the nucleus. Apart from the nucleus, mitochondria are the only organelles that contain DNA, which encodes around ten genes. This means that SCNT-ES cells might activate the immune system of an individual who is ostensibly being treated with their 'own' SCNT-ES cells and cause them to be rejected.

On the other hand, it makes these cells suitable for studying mitochondrial diseases, which are maternally inherited. However, to create their SCNT-ES cells, Tachibana *et al.* used nuclei from the cells of a patient with Leigh

syndrome — a disorder that can be caused by mutations in a mitochondrial gene. Because mitochondria in SCNT-ES cells originate from the oocyte, these cells would not carry the same mutation, nor model Leigh syndrome. Nonetheless, the cells generated are a proof of principle that adult somatic cells can be used in human SCNT.

The present study shows that the authors' SCNT-ES cells meet the main criteria for pluripotency: they can differentiate *in vitro*; they express pluripotency genes; and, when injected into an immune-deficient mouse, they form a teratoma — a type of tumour that contains many different cell types. Nevertheless, other properties of these cells were not extensively explored.

For instance, although human iPS cells are known to accumulate mutations without the necessary care⁷, overall they are very similar to ES cells derived from normal 'surplus' human embryos obtained by *in vitro* fertilization (IVF) treatment, and they are genetically and epigenetically stable under careful culture conditions over long periods⁸. Tachibana *et al.*, however, did not compare the efficiency of SCNT-ES cells, human IVF ES cells and iPS cells at differentiating *in vitro* under optimal conditions. Similarly, it is unclear whether SCNT-ES cells remain stable over time. Further investigation along these lines would be beneficial.

What this study provides is an excellent source of reference. Direct reprogramming of human iPS cells takes several weeks, whereas SCNT-ES cells are reprogrammed within a few hours by the natural factors present in the oocyte, and could in principle give rise to new offspring. A head-to-head comparison of these cell types over a long culture period would be ideal, not least to identify factors that might improve the efficiency and yield of direct reprogramming.

A cautionary note: since the paper's publication, there has been ongoing discussion about some errors, such as possible figure duplication and mislabelling, that it contains⁹. We therefore eagerly await experimental confirmation of Tachibana and co-authors' results by others, as well as the outcome of an investigation by *Cell* to determine how such errors occurred and whether they affect the study's overall conclusions (as *Nature* went to press, the results of this investigation had not been released). These concerns notwithstanding, the present findings are a major development, particularly for those studying human reproduction and IVF. Whether it is a game changer for research into understanding disease, regenerative medicine and drug discovery is debatable. ■ [SEE COMMENT P.159](#)

Christine L. Mummery is in the Department of Anatomy and Embryology, Leiden University Medical Centre, 2300 RC Leiden, the Netherlands. **Bernard A. J. Roelen** is

in the Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, 3584 CM Utrecht, the Netherlands. e-mails: c.l.mummery@lumc.nl; b.a.j.roelen@uu.nl

1. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. *Nature* **385**, 810–813 (1997).
2. Munsie, M. J. *et al. Curr. Biol.* **10**, 989–992 (2000).
3. Wakayama, T. *et al. Science* **292**, 740–743 (2001).

4. Tachibana, M. *et al. Cell* <http://dx.doi.org/10.1016/j.cell.2013.05.006> (2013).
5. Cyranoski, D. *Nature* **439**, 122–123 (2006).
6. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
7. The International Stem Cell Initiative *Nature Biotechnol.* **29**, 1132–1144 (2011).
8. Yamanaka, S. *Cell Stem Cell* **10**, 678–684 (2012).
9. Cyranoski, D. & Check Hayden, E. *Nature* <http://dx.doi.org/10.1038/nature.2013.13060> (2013).

CONDENSED-MATTER PHYSICS

Spintronics, the atomic way

Ultracold atomic gases are excellent platforms for exploring phenomena in condensed-matter physics. They have now been used to engineer the spin Hall effect and to make the atomic counterpart of the spin transistor. [SEE LETTER P.201](#)

PETER VAN DER STRATEN

The spin of elementary particles is a concept in quantum mechanics that has no counterpart in the classical world. Although the electron's spin was inferred from the 1922 Stern–Gerlach experiment on the deflection of particles, it has only recently been exploited in electronic devices — conventional electronic circuitry is based on the electron's charge. Atoms also carry a spin, and so could similarly be exploited for spin-based electronics. On page 201 of this issue, Beeler *et al.*¹ demonstrate how ultracold atoms can be used to build the atomic analogue of an electronic switch that was proposed more than 20 years ago: the spin transistor².

Spin-based electronics, or spintronics, has seen tremendous developments in the past decade³. At the heart of this field is the manipulation of the electron's spin, which in condensed-matter materials can be rather complicated owing to the interaction of this spin with its surroundings and the limited

controllability of the materials. In this regard, an effect called spin–orbit coupling, which allows the manipulation of the electron's spin without the use of local magnetic fields, has been instrumental in simplifying the construction of spintronic devices.

During the past ten years, ultracold atomic gases have become the ideal playground in which to investigate fundamental phenomena in many fields of physics, particularly condensed-matter physics. The excellent manipulation of both the internal and the external degrees of freedom of ultracold atoms allows the study of complex physics in a controlled way. Atoms have the advantage that they carry spin but have no charge. Therefore, charge effects that would otherwise need to be considered can be excluded. Furthermore, atoms can have either a fermionic or a bosonic particle character (they have half-integer or integer spin, respectively), so the effect of different quantum-particle statistics on spin interactions can readily be tested. Finally, ultracold atomic gases confer more versatility

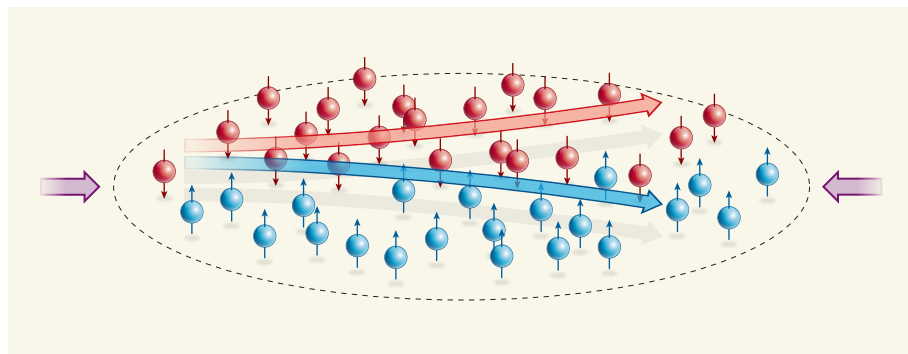


Figure 1 | Going separate ways. Beeler *et al.*¹ have demonstrated that two laser beams (purple arrows) can be directed at a cloud of atoms to generate an artificial magnetic field (not shown) that makes atoms of two spin orientations (red and blue) move in opposite directions.

on spin manipulation and detection than do electrons, yielding new ways to explore the richness of spintronics.

To make spintronic devices based on atomic spins, spin-orbit coupling needs to be addressed. In atomic physics, the effect couples the electron's spin with its orbital angular momentum through the electrical field of the nucleus, and leads to the fine structure (small splitting) of the internal states of the atom. In condensed-matter systems, however, the coupling is between the electron's spin and its linear motion, and is caused by the electrical field of the underlying atomic lattice. Researchers have recently proposed ways to create artificial electromagnetic fields (gauge fields) to induce such spin-motion interaction in atomic systems. These fields are produced by coupling internal states of atoms with laser beams at ultra-low temperatures. In their experiment, Beeler *et al.* used one such field to engineer spin-motion interaction in an atomic system and to observe a quantum effect known as the spin Hall effect.

The spin Hall effect is similar to the conventional Hall effect, in which the positively and negatively charged particles of an electrical conductor are separated by a transverse magnetic field, producing a voltage at a right angle to the current. In the spin Hall effect, the particles go their separate ways according to whether their spin points in one of two opposing directions. By using their gauge field, which was created by means of two laser beams, Beeler *et al.* convincingly show that atoms with opposite spin states that travel at right angles to the magnetic field produced by the gauge field move in opposite directions (Fig. 1). The researchers go on to show that their data agree well with theoretical calculations, indicating that the authors have a proper understanding of the mechanism behind the spin Hall effect in their system.

But Beeler and colleagues have taken their work one step further: they realized the atomic analogue of a spin transistor. Although previously proposed² in 1990, the spin transistor has been made only within the past few years⁴. By using the displacement of the atoms in the gauge field as the transistor's voltage difference between the drain and source electrodes, and the strength of the two laser beams as the transistor's voltage of the gate electrode, Beeler *et al.* realized an atomic system that shows the characteristics of a typical spin transistor. The simplicity and robustness of the authors' transistor also makes it a good option for splitting atoms according to their spin in a device known as a Mach-Zehnder interferometric sensor.

Beeler and colleagues' experiment opens up many avenues in the field of ultracold atomic gases. The gauge field produced is of the 'Abelian' type; however, there are proposals to generate non-Abelian gauge fields⁵. These non-Abelian fields are more difficult to realize experimentally but allow a closer

comparison with condensed matter, in which non-Abelian fields are usually responsible for the spin-motion interaction. Beeler *et al.* induce the spin Hall effect in their system by using spin-motion coupling, but in condensed matter the effect can also be induced by scattering of electrons by impurities. Although ultracold atomic gases are free of impurities, interactions between the atoms can be tuned to be made strong and yield exotic phenomena such as superfluidity.

Spin-orbit interactions can lead to topological insulators, which are insulating in their bulk but have topologically protected conducting states on their boundaries. Such states can easily be controlled and detected in ultracold atomic gases. The crossroads between ultracold atomic gases and condensed-matter physics provide fertile ground for research:

STEM CELLS

Regulation by alternative splicing

Stem-cell differentiation is controlled by RNA processing — as well as by gene expression and transcription. This finding is a milestone towards realizing these cells' potential for research and therapy. [SEE LETTER P.241](#)

YAIR AARONSON & ERAN MESHORER

Mammalian genomes contain some 23,000 genes. Yet the process of alternative splicing ensures that the number of proteins arising from these genes is at least ten times greater¹. It achieves protein diversity by varying the way in which the RNA transcript of a gene is processed: each of the protein-coding sections of a transcript can be either spliced out or left in to form different mature messenger RNAs. Consequently, multiple variants of a protein (isoforms) can be produced from a single gene, in a tissue-specific or developmental-stage-specific manner. On page 241 of this issue, Han *et al.*² describe the role of alternative splicing in the regulation of embryonic stem cells, thereby adding another notable regulatory layer to the known mechanisms that govern stem-cell state and differentiation*.

Embryonic stem (ES) cells have two special qualities: they can undergo an unlimited number of divisions, and they are pluripotent — that is, they can differentiate into any cell type of a mature organism. These cells, therefore, have great potential for clinical use and can serve as models for studying disease. Research into how ES-cell pluripotency is regulated has

the former focuses on fundamental knowledge obtained through the study of well-characterized systems under controllable conditions, whereas the latter applies such knowledge in information technologies. Many new phenomena can be expected to surface in these areas in the next few years. ■

Peter van der Straten is at the Debye Institute for Nanomaterials Science, Utrecht University, 3508 TA Utrecht, the Netherlands.
e-mail: p.vanderstraten@uu.nl

1. Beeler, M. C. *et al.* *Nature* **498**, 201–204 (2013).
2. Datta, S. & Das, B. *Appl. Phys. Lett.* **56**, 665–667 (1990).
3. Insight: Spintronics *Nature Mater.* **11**, 367–416 (2012).
4. Koo, H. C. *et al.* *Science* **325**, 1515–1518 (2009).
5. Dalibard, J., Gerbier, G., Juzeliūnas, G. & Öhberg, P. *Rev. Mod. Phys.* **83**, 1523–1543 (2011).

mainly focused on the control of gene expression through modifications of chromatin³ (the complex of DNA and proteins in chromosomes) and during transcription⁴. But, despite several master regulators of alternative splicing having been identified^{5,6}, none has been implicated in ES-cell maintenance, differentiation or transcription (Fig. 1).

To screen for alternative-splicing events associated with pluripotency, Han *et al.* studied RNA data from pluripotent cells and various differentiated cells from humans and mice. The pluripotent cells they investigated included not only ES cells but also induced pluripotent stem (iPS) cells, which are ES-like cells derived through molecular reprogramming of differentiated cells⁷.

The authors identified dozens of alternative-splicing events that differed between pluripotent and differentiated cells, including a previously known⁸ ES-cell-specific event in the mRNA for the pluripotency factor FOXPI1. And when they measured the expression levels of many known splicing regulators, the authors found a few that differed significantly between pluripotent cells and differentiated cells. In particular, two of the regulators — MBNL1 and MBNL2 — showed very low expression in ES cells and much higher expression in differentiated cells.

How do MBNL1 and MBNL2 affect stem-

*This article and the paper under discussion² were published online on 5 June 2013.

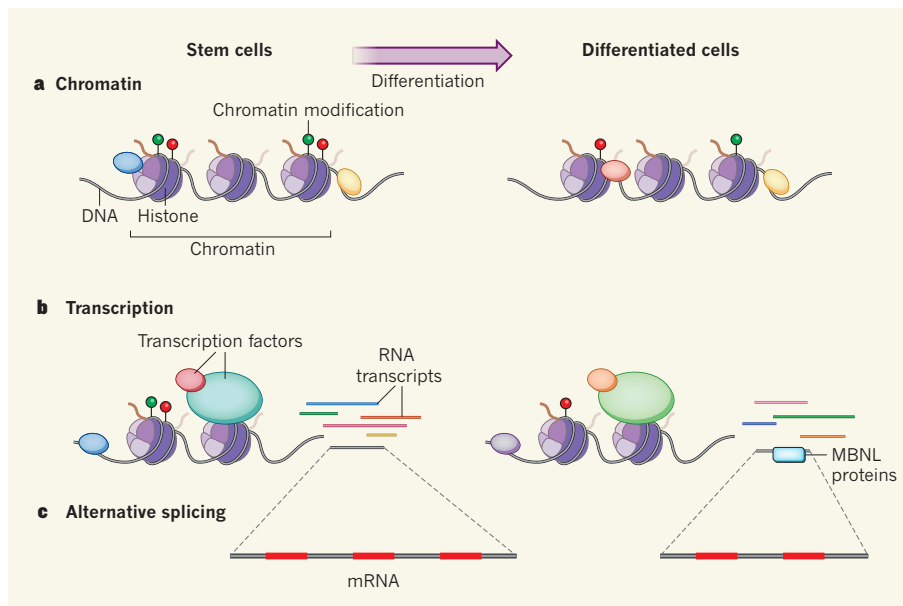


Figure 1 | Multilayer regulation of cell differentiation. **a**, The nature and amounts of chromatin-modifying and -remodelling proteins (blue and yellow) that bind to DNA and its associated histone proteins differ between stem cells and differentiated cells, affecting gene expression through regulation at the level of chromatin. **b**, At the level of transcription, specific transcription factors orchestrate the distinct transcript output of stem cells compared with differentiated cells. **c**, Han *et al.*² show that at the level of transcript processing, regulators of alternative splicing, such as MBNL proteins, govern the differences in mRNA, and thus protein, output between stem cells and differentiated cells.

cell identity? The researchers report that sites of alternative splicing in ES-cell transcripts are highly enriched in MBNL1- and MBNL2-binding motifs, and that these factors specifically bind to the sites in a unique pattern. So it seems that the binding patterns of these regulators control the omission or inclusion of protein-coding regions (exons) in the mature mRNA.

The cellular levels of MBNL proteins also seem to affect the differentiation state. Increased expression of these proteins in ES cells induced differentiation-specific alternative-splicing events, and decreased the levels of an ES-cell-specific isoform of FOXP1. Consistently, reducing expression of these proteins in differentiated cells led to a switch of the alternative-splicing program to an ES-cell-like pattern. And the efficiency of reprogramming of differentiated cells into iPS cells was greatly enhanced with reduced expression of MBNL1 and MBNL2 (the splicing pattern associated with 'stemness' was particularly prominent in cells that were successfully sustained through the later parts of the reprogramming process).

Han and co-workers' paper sets the stage for extensive follow-up studies. Understanding the exact mechanism of action of the MBNL proteins might help to identify upstream elements of this regulatory network. Moreover, the epigenetic state of ES cells — that is, genomic modifications that affect gene expression without changing the DNA sequence — is subject to continuous regulation, and a link between epigenetics and alternative splicing has been proposed^{9,10}. Understanding how

alternative splicing interacts with epigenetic and other networks that are known to regulate pluripotency would be fascinating. Furthermore, Han *et al.* identified many more sites of alternative splicing, and differential regulators of splicing in ES cells that they could not

investigate in the current work. These should be studied, as they might provide additional insights into the mechanism by which alternative splicing controls pluripotency.

The authors' observations might also have a notable practical implication. Splicing regulators could potentially be harnessed to control the efficiency and outcome of cellular differentiation and reprogramming — akin to the use of transcription factors for these purposes. While we tune in for follow-up studies, Han and colleagues' findings will surely change the ways in which researchers examine and manipulate pluripotent cells. ■

Yair Aaronson and Eran Meshorer are in the Department of Genetics, Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra (Givat Ram) Campus, Jerusalem 91904, Israel.
e-mail: meshorer@huji.ac.il

1. Nilsen, T. W. & Graveley, B. R. *Nature* **463**, 457–463 (2010).
2. Han, H. *et al.* *Nature* **498**, 241–245 (2013).
3. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. *Nature Rev. Mol. Cell Biol.* **12**, 36–47 (2011).
4. Ng, H.-H. & Surani, M. A. *Nature Cell Biol.* **13**, 490–496 (2011).
5. Zhang, C. *et al.* *Genes Dev.* **22**, 2550–2563 (2008).
6. Licatalosi, D. D. & Darnell, R. B. *Nature Rev. Genet.* **11**, 75–87 (2010).
7. Yamanaka, S. & Blau, H. M. *Nature* **465**, 704–712 (2010).
8. Gabut, M. *et al.* *Cell* **147**, 132–146 (2011).
9. Schwartz, S., Meshorer, E. & Ast, G. *Nature Struct. Mol. Biol.* **16**, 990–995 (2009).
10. Tilgner, H. *et al.* *Nature Struct. Mol. Biol.* **16**, 996–1001 (2009).

MALARIA

The vector as protector

Malaria infections are not always lethal. One reason for this may be that transmission from mosquitoes creates malaria parasites that trigger a more protective mammalian immune response. [SEE LETTER P.228](#)

ANDREW F. READ & NICOLE MIDEO

Malaria parasites can kill people, but death is not inevitable. Most infected individuals recover, some after experiencing relatively mild symptoms or none at all. What accounts for this variability? Host factors such as the expression of sickle-cell genes or acquired immunity are part of the explanation. But it is also well known that malaria parasites themselves can be more or less nasty^{1,2}. In this issue, Spence *et al.*³ (page 228) report a set of clever experiments in a mouse model of malaria infection that shows that the conditions experienced by parasites before they reach the mammalian bloodstream can

determine just how virulent they are*.

Malaria parasites transmitted to people by mosquitoes migrate to the liver, where they replicate before entering the bloodstream. For convenience, and because only blood-stage parasites cause disease, most experimental studies of malaria in humans and animals bypass the mosquito and liver stages and inject parasites directly into the bloodstream. Using the malaria parasite *Plasmodium chabaudi*, which infects rodents, Spence and colleagues compared the blood-stage infections generated by this method with those initiated naturally, by mosquito bite. They found that, compared with directly injected parasites,

*This article and the paper under discussion³ were published online on 29 May 2013.

mosquito-transmitted parasites replicated less well once in the bloodstream and generated lower-grade infections that persisted for longer. Moreover, these parasites did not induce the severe weight loss, hypothermia and liver damage caused by parasites injected directly into the bloodstream.

Why these differences? An important clue came from the authors' finding that, in immunodeficient mice, parasites transmitted by mosquitoes grew just as well as those injected directly. This suggested that there is nothing intrinsically attenuated about parasites derived from mosquitoes. Spence *et al.* show that mosquito-transmitted parasites elicit a qualitatively different immune response in the mouse — one that better controls parasite replication and relies less on the inflammatory signalling molecules that are associated with severe disease. To try to explain this difference, Spence *et al.* conducted a genome-wide RNA analysis and found that mosquito transmission modifies the expression of about 10% of the genome of blood-stage parasites. Intriguingly, expression was most intensely regulated for gene families encoding antigenic proteins, against which the host's immune system mounts its response. The hypothesis, then, is that mosquito transmission alters subsequent antigen expression when the parasites are in the bloodstream, and that the induced gene-expression pattern elicits an immune response that more effectively contains the parasites with less collateral damage to the host.

It seems that it is the environment experienced by the parasites during natural transmission that triggers this 'attenuated phenotype'. That environment could be inside the mosquito itself, or it could be something experienced by the parasite in the skin soon after injection, during its journey to the liver or in the liver. Intriguingly, Spence *et al.* show that the attenuated phenotype also occurs in mice injected with blood-stage parasites isolated from other mice with mosquito-initiated infections. Thus, the phenotype is stable for several cycles of blood-stage parasite replication, although it does gradually decay over subsequent rounds of injecting these parasites into new hosts. It will be interesting to determine whether profiles of the host immune response and of parasite-antigen expression associated with attenuation decay in a similar manner.

Does this discovery mean that all future experimental malaria infections should be initiated by mosquitoes? There is no way to include mosquito transmission in *in vitro*

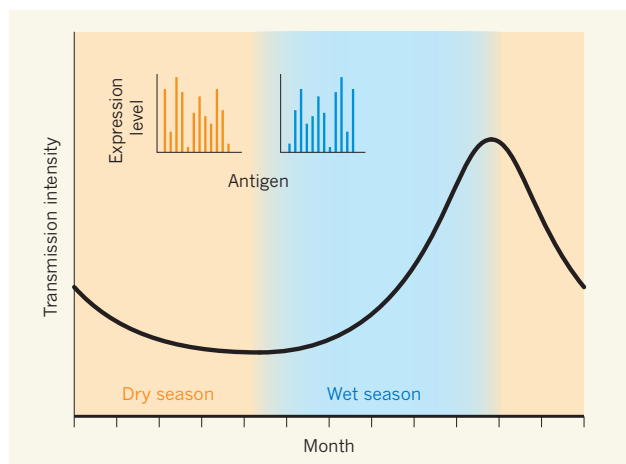


Figure 1 | Evolutionary selection of antigenic profiles. If variation in the antigens expressed by a parasite gives rise to qualitatively different infection dynamics and outcomes, natural selection might favour the expression of different antigenic profiles at different times or in different regions. For instance, to survive a prolonged dry season, when little to no transmission occurs, parasites with the attenuated phenotype described by Spence *et al.*³ — causing chronic infections of low virulence — and the associated antigenic profile may be most successful. By contrast, when the rainy season begins and epidemic situations arise, parasites with antigen-expression profiles that result in rapid proliferation and transmission may be favoured. In this case, the cost of shorter infectious periods associated with rapid clearance of the parasite by the immune system, or host death, may be offset by the advantages of faster transmission to new hosts. These evolutionary forces might generate parasites that respond to cues associated with transmission (through altered gene expression) in some regions, and parasites that do not in others, such as in endemic areas where transmission occurs year-round.

studies of the most lethal human malaria parasite, *Plasmodium falciparum*. Immunosuppressed mice with human-cell transplants can support *P. falciparum* infections⁴, but it is unclear whether the addition of one aspect of biological reality (mosquito transmission) will make up for the loss of another (the use of human parasites in mice). Mosquito infections are an option in animal models, from which much has already been learned by injecting blood-stage parasites. For example, experiments with *P. chabaudi* have shown that a powerful contributor to the severity of malaria can be the host immune response itself⁵, and that competition between different parasite strains can be a potent force shaping the evolution of drug resistance⁶. The key question is not whether these phenomena still occur if more of the parasite life cycle is incorporated into the experimental work, but whether they occur in nature.

The effects of mosquito transmission on host immune response and parasite antigen expression observed by Spence and colleagues might be the independent outcomes of environmental influences, or they might be causally connected. If the latter is true, the question remains whether immunity triggers the antigenic profile or whether altered antigen expression triggers a more protective immune response. The direction of this causality could have implications for vaccine design.

In the first scenario, considering the characteristics of the immune response generated by a vaccine would be important for protecting against severe disease if vaccination does not completely block infection. In the second case, a vaccine that results in exposure to a particular antigenic profile may be crucial for developing an optimally protective response. Spence *et al.* compared infections initiated by mosquitoes and by blood-stage parasites at just one time in the blood-stage infection, but antigen expression can be highly variable in time and across host tissues^{7,8}, so further assessment of these profiles is needed.

If antigen-expression profiles are indeed a major determinant of malaria-parasite virulence, and if these are not completely constrained by the parasite's developmental requirements, we predict that natural selection will favour different antigen-expression profiles in different epidemiological settings (Fig. 1). If this is the case, then virulence variability due to genetic polymorphisms or phenotypic plasticity will be common in nature. This might explain apparently contrasting experimental results. For example, Spence *et al.* found that mosquito transmission

attenuated parasite replication in two clones of *P. chabaudi*, but earlier experiments using a different clone found no such effect⁹. Similarly, physicians who deliberately infected people with *P. falciparum* to treat neurosyphilis reported the same clinical picture regardless of how the infection was initiated². Clearly, much is yet to be learned about how malaria parasites make people sick, and about the role of the mosquito vector in modulating the disease it initiates. ■

Andrew F. Read and Nicole Mideo are at the Center for Infectious Disease Dynamics, Departments of Biology and Entomology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.
e-mail: a.read@psu.edu

1. Mackinnon, M. J. & Marsh, K. *Science* **328**, 866–871 (2010).
2. Covell, G. & Nicol, W. D. *Br. Med. Bull.* **8**, 51–58 (1951).
3. Spence, P. J. *et al. Nature* **498**, 228–231 (2013).
4. Legrand, N. *et al. Cell Host Microbe* **6**, 5–9 (2009).
5. Langhorne, J., Ndungu, F. M., Sponaas, A.-M. & Marsh, K. *Nature Immunol.* **9**, 725–732 (2008).
6. Read, A. F., Day, T. & Huijben, S. *Proc. Natl Acad. Sci. USA* **108**, 10871–10877 (2011).
7. Ebbinghaus, P. & Krücken, J. *Malaria J.* **10**, 272 (2011).
8. Lawton, J. *et al. BMC Genomics* **13**, 125 (2012).
9. Mackinnon, M. J., Bell, A. & Read, A. F. *Int. J. Parasitol.* **35**, 145–153 (2005).

COSMOLOGY

Hydrogen wisps reveal dark energy

Traces of hydrogen gas, detected over vast regions of space, have for the first time been used as a standard ruler to measure dark energy — the unknown cosmic energy that is causing the Universe's expansion to speed up.

TAMARA M. DAVIS

Writing in *Astronomy & Astrophysics* and the *Journal of Cosmology and Astroparticle Physics*, the Baryon Oscillation Spectroscopic Survey (BOSS) collaboration reports^{1,2} on how it has used observations of cosmic hydrogen gas to determine the expansion rate of the Universe back before the epoch of acceleration began.

That the expansion of the Universe began to accelerate about 7 billion years ago is now an accepted part of the standard cosmological model. Yet what is causing the acceleration remains a mystery. The term 'dark energy' encompasses several theoretical possibilities, but to distinguish between them more information is needed, such as whether dark energy is constant or changes over time.

So far, astronomers have been limited by the fact that most observations have targeted the relatively nearby Universe (at redshift (z) of less than about 1, where supernovae and galaxies are easy to see) or far away (at z about 1,100, where the remnant radiation from the Big Bang is seen). The BOSS team bridges this redshift gap, and measures cosmic expansion at z about 2.3, corresponding to a time when the Universe was less than a quarter of its present estimated age of 13.8 billion years.

This stunning measurement confirms the existence of dark energy and, most interestingly, shows no sign that it has varied over the past 10 billion years. Dark energy remains consistent with Einstein's cosmological constant — a result that could easily have been disproved with data from these distances. This measurement was made possible by the efforts of the BOSS team. The collaboration is in the process of obtaining the spectra of 1.6 million galaxies and 150,000 quasars (the extremely luminous central parts of active galaxies) using the Sloan Digital Sky Survey 2.5-metre telescope at Apache Point Observatory in Sunspot, New Mexico. The purpose is to determine the distribution of matter across more than half of the observable Universe. The distribution is not random, and it holds a wealth of information about dark energy, dark matter and the strength of gravity.

Ever since it was revealed by supernovae observations in 1998 that the expansion of

the Universe is accelerating, enormous effort has been put into measuring this acceleration in enough detail to try to elucidate the cause. Among the primary observable parameters have been baryon acoustic oscillations. These oscillations were formed by sound waves in the early Universe, which was then so dense that sound travelled everywhere at more than half the speed of light. About 300,000 years after the Big Bang, the Universe had expanded to the extent that matter was no longer dense enough for sound waves to propagate. The waves then froze into place, leaving a characteristic scale imprinted on the primordial density distribution from which galaxies would eventually form.

Over the past decade, galaxy surveys such as the two-degree Field Galaxy Redshift Survey³, the six-degree Field Galaxy Survey⁴, the Sloan Digital Sky Survey⁵ and the WiggleZ Dark Energy Survey⁶ have mapped the distribution of galaxies in the Universe at ever increasing distances. They have all revealed the characteristic baryon-acoustic-oscillation scale in the distribution pattern of galaxies. Using that scale as a standard ruler, the acceleration of the expansion has been beautifully confirmed with a precision that now equals that obtained by supernovae studies.

BOSS is the next survey in that distinguished

line, but the team has seen even farther by observing quasars. The extreme brightness of quasars arises because of the heating of material actively falling into the central black holes of the galaxies that host them. The team used these quasars as 'backlights' to detect the wispy hydrogen gas that permeates the Universe between the galaxies. When sunlight passes through the branches of a tree, the pattern of leaves can be inferred by the shadows that they cast (Fig. 1). Similarly, when quasar light passes through hydrogen clouds, the absorbed light gives a map of where the hydrogen lies.

This is not the first time that hydrogen has been traced using quasars as backlights. Nor is it the first time that baryon acoustic oscillations have been measured. However, it is the first time that hydrogen has been used to measure baryon acoustic oscillations, and it is by far the most distant measurement of the Universe's expansion rate so far.

This spectacular result supports the idea that the simplest model of dark energy — that it is constant — really is the best one. It leaves astronomers in an interesting position. Just as particle physicists found the Higgs boson exactly where they expected to find it, cosmologists have found dark energy exactly where the simplest theory predicts it to be.

Where do we go from here? The problem remains that there is no good theoretical explanation for either dark energy or dark matter. Observations will continue to improve, but it is becoming clear that the real breakthroughs needed are theoretical. In a sense, observations are easy. Given more time, better equipment, careful analysis and increased person-power, measurements can always be improved. Theory is much more difficult. No amount of time will guarantee a breakthrough, and it is possible that the 'next big thing' could become lost in the morass of poor theories for lack of a sufficiently charismatic proponent.

In the next generation of cosmology



Figure 1 | Useful shadows. From the pattern of shadows on the ground, one can infer the pattern of the leaves of a tree. The BOSS team^{1,2} has used the 'shadows' in spectra of distant quasars to infer the distribution of hydrogen gas in the Universe, and to detect the effects of dark energy.

MASKOT/GETTY

experiments, observers, aware of this issue, are doing their best to inform theory in different ways. They are providing different types of observations that could assist in distinguishing between theories and in directing theorists' investigations.

Meanwhile, I continue to be awed by the fact that humans are able to measure the distribution of hydrogen as it was more than 6 billion years before Earth formed, and to relate it to sound waves in the infant Universe

by applying only simple physical concepts, such as pressure and gravity, which also govern everyday life on Earth. That fact further increases my confidence in the overall picture that cosmology has revealed, and is inspiration enough to continue efforts to figure out the remaining mysteries. ■

Tamara M. Davis is in the School of Mathematics and Physics, University of Queensland, Brisbane 4072, Australia.

e-mail: tamarad@physics.uq.edu.au

1. Busca, N. G. *et al. Astron. Astrophys.* **552**, A96 (2013).
2. Slosar, A. *et al. J. Cosmol. Astropart. Phys.* 026; <http://dx.doi.org/10.1088/1475-7516/2013/04/026> (April 2013).
3. Eisenstein, D. *et al. Astrophys. J.* **633**, 560–574 (2005).
4. Beutler, F. *et al. Mon. Not. R. Astron. Soc.* **416**, 3017–3032 (2011).
5. Anderson, L. *et al. Mon. Not. R. Astron. Soc.* **427**, 3435–3467 (2012).
6. Blake, C. *et al. Mon. Not. R. Astron. Soc.* **418**, 1707–1724 (2011).

soil-carbon sequestration and global warming, compared with those of conventional tillage, have been studied and debated extensively. Three points are generally accepted. First, when conventional tillage is replaced by no-tillage cultivation, soil carbon increases in the soil surface layers⁶, but few significant differences in soil carbon are found in deeper layers⁷. Second, nitrous oxide emissions might increase in recently converted systems, but are lower in the long term^{6,8}. And third, emissions of the greenhouse gas methane remain essentially the same⁶.

Similarly, converting conventionally tilled systems to grasslands (as in the CRP) generally leads to increases in soil carbon^{1,4}. It is also expected to decrease nitrous oxide emissions and increase the uptake of methane into the soil^{5,9}, leading to a general decrease in global-warming impact (Fig. 1b). But no study has investigated the effect of converting restored CRP grassland to arable land on greenhouse-gas emissions and global warming — until now.

Ruan and Robertson measured the main soil-derived greenhouse gases (carbon dioxide, nitrous oxide and methane) in four fields that had been managed under the CRP. Three of those fields were converted to soya-bean cultivation and divided into no-tillage and conventional-tillage plots; the fourth field was maintained as grassland. The authors also took ancillary measurements, such as soil temperature, moisture, density and mineral nitrogen content, all of which could potentially explain any observed differences in greenhouse-gas emissions between the fields.

A limitation of the study is that the researchers took measurements for only one year. But such short-term assessments are crucial, because the greatest changes in greenhouse-gas emissions are expected shortly after grassland is ploughed up, and in the first year of cultivation as the system responds to the change — often undergoing extreme, but possibly only transitory, changes in soil structure, soil nutrients and plant growth^{6,10}.

Ruan and Robertson observed

CONSERVATION

Spare our restored soil

The conversion of poor-quality arable lands to grassland has prevented soil erosion and sequestered carbon. A study finds that greenhouse gases will be emitted if these lands return to cultivation, especially if they are ploughed.

JOHAN SIX

In 1985, the US Department of Agriculture established the Conservation Reserve Program to revitalize degraded and marginal agricultural land by converting it to grassland. Although enthusiasm for the programme has been far from universal, many agree that it did generally rebuild soils and cause carbon to be sequestered within them¹. But with a large number of the programme's contracts with farmers about to expire, revitalized land might soon be ploughed up, with unknown environmental consequences. Writing in *Global Change Biology*, Ruan and Robertson² fill us in on one of those consequences: the effect on greenhouse-gas emissions. They conclude that it would be best to maintain these restored grasslands as they are, but that, if they must be cultivated, 'no-tillage' farming produces many fewer greenhouse-gas emissions than does farming involving conventional ploughing.

Ploughing up land results in the depletion of soil carbon — on average, about 50% of the carbon is lost, compared with the amount maintained under naturally occurring vegetation^{3,4}. Emissions of the greenhouse gas nitrous oxide are also drastically increased because of fertilizer use and increased nitrogen mineralization in soil when land is cultivated⁵. The global-warming impact is thus drastically increased when land is ploughed up for cultivation (Fig. 1a). Another profound effect of ploughing and cultivating land is that it decreases the structural stability of soil, leading

to erosion. This became a major concern in the 1970s.

The Conservation Reserve Program (CRP) was put in place to 'retire' vulnerable lands by restoring them to grassland, with the goal of preserving soils and all the services they can provide, such as storing water, carbon and nutrients. At the same time, no-tillage practices — made possible by the availability of herbicides — were promoted to conserve arable soils. More recently, no-tillage cultivation has been advocated as a potential tool for mitigating or adapting to climate change.

The effects of no-tillage practices on

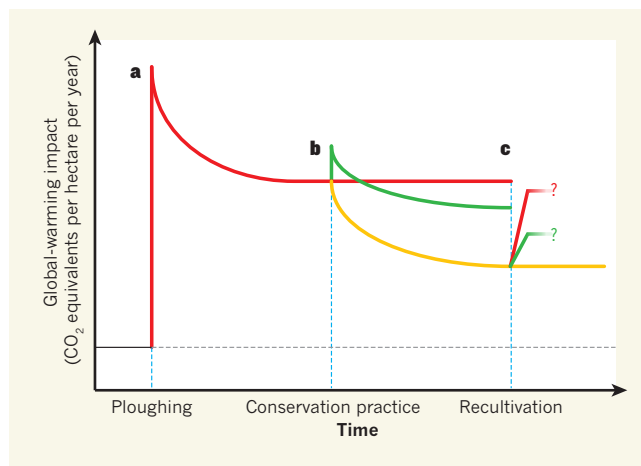


Figure 1 | Land-use change influences greenhouse-gas fluxes from soils to the atmosphere. a, Ploughing grasslands to create arable land increases greenhouse-gas emissions (red line) compared with those produced by the natural ecosystem (grey broken line), especially in the period shortly after ploughing, until a new equilibrium is reached. b, Emissions can be reduced, especially in the long term, by restoring arable land to grassland (yellow line), or, less effectively, by using no-tillage practices (green line) on the arable land. c, Ruan and Robertson² report that, if restored grassland is recultivated using conventional tillage, greenhouse-gas emissions increase in the year after ploughing; the effect is smaller if no-tillage practices are used. The amount of emissions in the longer term remains unknown.

that CRP-managed grasslands are a net greenhouse-gas sink, whereas soya-bean agro-ecosystems are a net source of greenhouse gases. The authors' findings confirm that we will lose some of the environmental services provided by CRP land if we cultivate it. However, they also report that soya-bean systems under no-tillage management have less than half of the global-warming impact of those that are conventionally tilled (Fig. 1c).

As is usual for agricultural systems, nitrous oxide emissions accounted for most of the differences in global-warming impact, emphasizing that this is the greenhouse gas we need to monitor and manage in such systems. This finding also begs the question: what would have happened if the CRP grassland had been converted to fertilized maize (corn) instead of unfertilized soya bean? The nitrous oxide emissions might have been even greater. But, if the CRP grassland had contained leguminous vegetation, as some CRP grasslands do, the differences between the cultivated and non-cultivated systems might have been smaller. However, the most pertinent remaining question is, what long-term effect will cultivation of CRP land have on greenhouse-gas emissions? Are the differences observed by Ruan and Robertson transitory, or will they become even bigger over time?

Unresolved issues aside, the authors' study does have one clear message: let us not lose the environmental services that have been provided by the CRP. No-tillage practices should be considered to attenuate the greenhouse-gas costs of ploughing up CRP land, but the best option for the environment is to maintain the land under grasses. Incentives for both options need to be provided. ■

Johan Six is in the Department of Environmental Systems Science, ETH-Zurich, 8092 Zurich, Switzerland.
e-mail: jsix@ethz.ch

1. Baer, S. G., Meyer, C. K., Bach, E. M., Klopff, R. P. & Six, J. *Ecosphere* **1**, art5 (2010).
2. Ruan, L. & Robertson, G. P. *Glob. Change Biol.* <http://dx.doi.org/10.1111/gcb.12216> (2013).
3. Haas, H. J., Evans, C. J. & Miles, E. F. *Tech. Bull.* No. 1164 (US Dpt Agriculture, 1957).
4. Ogle, S. M., Breidt, F. J. & Paustian, K. *Biogeochemistry* **72**, 87–121 (2005).
5. Mosier, A., Schimel, D., Valentine, D., Bronson, K. & Parton, W. *Nature* **350**, 330–332 (1991).
6. Six, J. et al. *Glob. Change Biol.* **10**, 155–160 (2004).
7. Kravchenko, A. N. & Robertson, G. P. *Soil Sci. Soc. Am. J.* **75**, 235–240 (2011).
8. Van Kessel, C. et al. *Glob. Change Biol.* **19**, 33–44 (2013).
9. Keller, M., Veldkamp, E., Weitz, A. M. & Reiners, W. A. *Nature* **365**, 244–246 (1993).
10. Mummey, D. L., Smith, J. L. & Bluhm, G. *Agric. Ecosyst. Environ.* **70**, 79–87 (1998).

EARTH SCIENCE

Water may be a damp squib

Experiments on silicon diffusion in the mineral olivine cast doubt on the widely held belief that water has a significant effect on the rheological properties of Earth's upper mantle. **SEE LETTER P.213**

JOHN BRODHOLT

In 1965, Griggs and Blacic demonstrated¹ that wet quartz is much weaker than dry quartz, and suggested that this should also be the case for other silicates. Indeed, the authors concluded: "These observations raise the possibility of great weakness in the earth's deeper crust and outer mantle at temperatures far below the melting point." In other words, water might control the viscosity of all rocks. Since then, experiments on other silicates have reinforced that view², and small amounts of water bound into normally dry minerals such as olivine are now attributed with almost divine powers in shaping the way that Earth works. On page 213 of this issue, Fei and colleagues³ show that water may be much less important in controlling many large-scale processes occurring in the Earth than was previously thought.

The ability of small amounts of water (or, more strictly, the concentration of hydroxide, OH⁻) to weaken minerals and rocks is known as hydrolytic weakening, and is implicated in a wide range of Earth and planetary processes. Indeed, it is argued that plate tectonics itself may owe its existence to hydrolytic weakening⁴. The tectonic difference between Venus and Earth may be because Earth has kept more of its water⁵. Moreover, Earth could be subducting more water into its mantle than it loses from volcanoes at mid-ocean ridges and hotspots, thereby further weakening the mantle and enhancing mantle convection rates, despite the overall background cooling⁶. The effect of water on the viscous strength of rocks is conveniently called upon when normal rock-deformation processes cannot be invoked as an explanation, and phrases such as 'water weakens rocks by orders of magnitude'



50 Years Ago

'The scientific revolution and leisure'—Many of our obligations eat into our spare time and, in a large city, the amount of spare time which is actually free may be very small. First of all there is the journey to and from work. Moreover, when the worker arrives home a queue of domestic duties may await him: shopping and the payment of bills, repairs in the home, the care of children, and perhaps a visit to the doctor or dentist ... preparations for a holiday may involve a gigantic effort ... All the same, it remains true that the free time enjoyed by the average worker has enormously increased during the past century. A hundred years ago the average expectation of life at birth was 40 years and a man worked about 70 hours a week. To-day, these figures are reversed, the expectation of life is 70 years and the working week is nearer 40 hours. Although there is, relatively speaking, plenty of free time, happiness, the Holy Grail of the twentieth century, remains as remote as ever.

From *Nature* 15 June 1963

100 Years Ago

It has been shown experimentally that fever is due to the digestion of proteins in the blood and in the tissues. Bacteria are living proteins. They get into the body and grow, converting the proteins of man's body into bacterial proteins. After a period of incubation the cells of the body pour out a ferment which digests and destroys the bacteria. In this process fever originates. In itself fever is beneficial; it is a manifestation of the attempt on the part of nature to destroy the invading organism. However, nature may overdo the matter, and fever *per se* becomes dangerous when it goes much above 105°.

From *Nature* 12 June 1913

are commonplace in Earth science. Fei and colleagues' results throw a spanner in the works by suggesting that hydrolytic weakening in olivine is a much smaller effect than generally thought.

It needs to be said upfront that Fei *et al.* did not directly measure the viscosity of wet versus dry rocks. They approached the problem from a different angle by measuring silicon diffusion. This is because the viscosity of high-temperature rocks is often controlled by the most slowly diffusing atom, which is silicon in the case of olivine. In dry olivine, for instance, the activation energies for creep and for silicon diffusion are almost the same, suggesting that creep in olivine is controlled by silicon diffusion². Fei and colleagues found that, surprisingly, water increases silicon diffusion by less than a factor of ten over three or more orders of magnitude in water content. If the viscosity of olivine is controlled by silicon diffusion, then this effect is much less than required for the several orders of magnitude of weakening commonly cited in the literature.

So which theory is right? First of all, it could simply be that the deformation mechanism in olivine is not controlled by silicon diffusion at all, and that the similar activation energy of creep and diffusion is just a coincidence. Another possibility is that the deformation experiments showing a strong hydrolytic weakening effect² were performed under water-saturated conditions, possibly enhancing other deformation mechanisms (such as sliding on crystalline-grain boundaries) and thereby producing an artificially weakened rheology. However, the strain rates in the deformation experiments should then show a dependence on grain size — something that the authors of the studies took pains to point out is not observed. And finally, Fei and colleagues' diffusion experiments were performed on iron-free olivines; ferric iron and other ionic species may affect both diffusion and deformation.

However, support for a small hydrolytic-weakening effect was published earlier this year⁷. These authors used a newer deformation apparatus to measure the rheology of wet olivine up to a pressure of 7 gigapascals (equivalent to a depth of about 200 kilometres). First, they found that wet olivines were only about 1.5 times weaker than dry olivines, and second, they saw no measurable dependence of viscosity on water after the first few parts per million or so of water. In other words, a small amount of weakening occurred from a small concentration of water, after which the strength remained the same regardless of the water content. Certainly, the authors did not see the large dependence of viscosity on water content seen in the earlier deformation experiments².

So is it possible that the large effect of water was never really there in the deformation experiments² to begin with? Wet olivine is certainly weaker than dry olivine, but the deformation experiments on which marked

hydrolytic weakening is based were performed at relatively low pressures (less than 0.5 GPa). At these pressures, the solubility of water in olivine is low, restricting the range of water contents in individual studies and perhaps making it difficult to determine the exact dependence of viscosity on water concentration. However, it is worth noting that the amount of water in the latest deformation experiments⁷ is also restricted, and so the uncertainty in determining the dependence of viscosity on water content could be aimed at their — opposite — results too.

It is early days. Fei and colleagues' diffusion experiments need to be repeated and extended to other compositions — particularly iron-bearing olivines. Deformation experiments need to be performed on olivines containing more water, and on polycrystalline material as well as single crystals. Deformation and diffusion mechanisms appropriate to Earth conditions and compositions need to be worked out. And, of course, Earth is not only olivine, so what about the other minerals of which it consists, such as pyroxenes, garnet, wadsleyite, ringwoodite and perovskite? How does their strength depend on water?

STRUCTURAL BIOLOGY

Ion channel twists to open

GIRK channels allow potassium ions to cross the cell membrane, thereby affecting the electrical status of the cell and so its functioning. Structural data now provide insight into the channels' mode of operation. [SEE ARTICLE P.190](#)

EITAN REUVENY

Ion channels are the main units responsible for the electrical activity in our body. They constitute a large family of some 400 proteins in humans. A subfamily of these proteins consists of four GIRK channels¹, which specialize in converting chemical signals — mostly those of neurotransmitter molecules such as acetylcholine, dopamine, serotonin and adrenaline — into electrical ones in heart cells and neurons. They are therefore essential for controlling heart rate and the activity of neural circuits. In this issue, Whorton and MacKinnon² (page 190) describe the long-awaited crystal structure of the mammalian GIRK2 channel in complex with two subunits of a G protein (a dimer of the G β and G γ subunits), providing information about their mechanism of opening*.

Activation of GIRK channels often begins with stimulation of G-protein-coupled

*This article and the paper under discussion² were published online on 5 June 2013.

Finally, what about all those Earth processes that seem to require hydrolytic weakening? It is worth pointing out that there are other ways of softening minerals and rocks. Melts, strain localization, grain-size reduction and changes in deformation mechanism can all produce interesting dynamical behaviour⁸. So, regardless of whether hydrolytic weakening is or is not a strong effect, plate tectonics exists, and Venus is definitely different from Earth. ■

John Brodholt is in the Department of Earth Sciences, University College London, London WC1E 6BT, UK.

e-mail: j.brodholt@ucl.ac.uk

1. Griggs, D. T. & Blacic, J. D. *Science* **147**, 292–295 (1965).
2. Kohlstedt, D. L. *Rev. Miner. Geochem.* **62**, 377–396 (2006).
3. Fei, H., Wiedenbeck, M., Yamazaki, D. & Katsura, T. *Nature* **498**, 213–215 (2013).
4. Regenauer-Lieb, K., Yuen, D. A. & Branlund, J. *Science* **294**, 578–580 (2001).
5. Moresi, L. & Solomatov, V. *Geophys. J. Int.* **133**, 669–682 (1998).
6. Korenaga, J. J. *Geophys. Res.* **116**, B12403 (2011).
7. Girard, J., Chen, J., Raterron, P. & Holyoke, C. W. *Phys. Earth Planet. Inter.* **216**, 12–20 (2013).
8. Bercowski, D. & Ricard, Y. *Phys. Earth Planet. Inter.* **202–203**, 27–55 (2012).

receptors (GPCRs) in the cell membrane (see Fig. 1a of the paper²). For instance, binding of acetylcholine to a muscarinic-type GPCR on a heart cell results in release of the G α and G $\beta\gamma$ subunits of the G protein that is attached to the GPCR at the intracellular surface of the cell membrane. G $\beta\gamma$ then activates GIRK channels^{3–5}, which allow efflux of intracellular potassium ions (K⁺) from the cell, causing hyperpolarization of the cell membrane (it becomes more negative inside relative to outside) and so reducing the cell's electrical excitability. Acetylcholine thus slows the heart rate.

Following the breakthrough discovery³ that the G $\beta\gamma$ dimer is responsible for opening GIRK channels after GPCR activation, extensive biochemical and electrophysiological studies have focused on the mechanism of activation of these channels and the role of associated modulatory molecules¹. These studies, however, fell short of deciphering the exact mode of interaction of the G $\beta\gamma$ subunits with

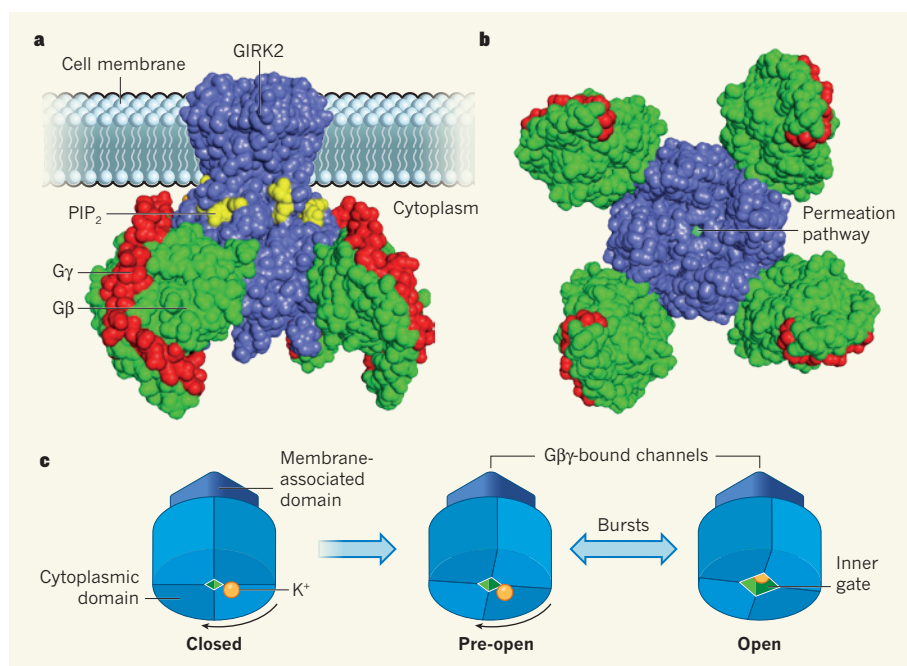


Figure 1 | The GIRK2 channel in action². Binding of the G β and G γ subunits of a G protein to each of the four GIRK2 monomers activates the homotetrameric GIRK2 channel. **a**, Side view of the channel, with the associated PIP₂ molecules and G β γ . **b**, Bottom view of the same complex. Note the four-fold symmetry and the centre permeation pathway for potassium ions (K⁺). **c**, Models of structural rearrangements associated with the opening of the GIRK2 channel by G β γ dimers. Looking from inside the cell, when G β γ binds the cytoplasmic-associated domain rotates clockwise relative to the membrane-associated domain to widen the permeation pathway at the inner gate. The resulting pre-open conformation, however, is not large enough to allow passage of hydrated K⁺ through the channel. Additional twisting in the same direction can widen the inner gate further, allowing hydrated K⁺ to move through this open conformation. Movement from the pre-open to the open conformation is random, and is thought to be the mechanism governing the well-known bursting activity of the channel.

the channel and the structural transitions that lead to channel opening.

Whorton and MacKinnon describe atomic-level interaction of the G β γ dimer with a GIRK channel consisting of four GIRK2 monomers (Fig. 1a,b). The 3.5-ångström-resolution structure shows that each of the four monomers is bound to a G β γ dimer, in agreement with previous biochemical evidence. They are also individually bound to a molecule of the phospholipid PIP₂ and a sodium ion, both of which are necessary for channel functionality.

The G β γ dimer and the channel share a relatively small surface area of contact (roughly 700 Å²), compared with the footprint of other known G β γ interactor molecules such as G α , the GPCR kinase-2, phospholipase-C β and phosphatidylinositol-3-OH kinase. Nevertheless, the contact areas of each interactor with G β γ overlap to various degrees, such that G β γ cannot bind to more than one interactor simultaneously; this underlines the singularity of the G β γ -mediated signalling event. The G β γ dimer seems to interact with the channel at the interface of the channel monomers, including regions that are known to be involved in channel activation, such as the LM loop. The interaction involves both short-range intermolecular forces such as van der Waals forces and hydrogen bonding and long-range electrostatic forces, and is further stabilized by anchoring of G β γ to the

cell membrane through the lipid moiety in the G γ subunit.

To reveal the structural changes associated with channel activation, Whorton and MacKinnon aligned three structures of GIRK2-PIP₂: the normal channel, the channel in complex with G β γ dimers and an always-active channel mutant. A comparison of the first two structures revealed two main conformational differences. On G β γ -dimer binding, there was a clockwise (looking from inside the cell) rigid-body rotation of about 4° along the centre axis of the channel, relative to the transmembrane domains. There was also a widening of the bottom of the channel's inner helices — the inner helical gate — on the cytoplasmic side.

The conformational changes that open the inner helical gate are comparable to the widening of a lens aperture by hand-rotating the aperture ring. In the resulting conformation, however, the gate is too narrow to allow hydrated K⁺ to pass through the channel. So how do G β γ dimers 'gate' the channel? Adding the always-active channel to the analysis provided an answer. In this structure, the rotation of the cytoplasmic domain relative to the transmembrane domain was more pronounced, causing the inner gate to widen further and permit the passage of hydrated K⁺.

On the basis of their observations, the

authors formulate a G β γ -dependent gating scheme for GIRK channels (Fig. 1c). Following activation of GPCRs and dissociation of the G protein into G β and G α , the free G β γ dimer diffuses to the inner membrane surface and binds to the channel molecule to induce a 'pre-open' state. In this state, rotation of the channel's cytoplasmic domain relative to its transmembrane domain broadens the inner helical gate, although the channel cannot conduct K⁺. Nevertheless, the pre-open conformation brings the channel to a higher energetic state, allowing it to make frequent random changes to the open conformation, by which K⁺ conduction can occur. Such frequent conformational changes are consistent with the well-characterized 'bursting' behaviour of the channel that is seen during recordings of single-channel activity.

Whorton and MacKinnon's data relate to homomeric GIRK2 channels, which are present only in selected areas of the brain⁶. And at least one other GIRK channel, GIRK1, differs from GIRK2 in the length of its amino-acid sequence and in several amino-acid residues involved in G β γ binding. So it is important to determine the structure of the two most prevalent channel species in the brain and heart — the GIRK1/GIRK2 and GIRK1/GIRK4 heteromers, respectively — in complex with the G β γ dimer. Although the overall structural transitions associated with gating are likely to be the same, the contact surface of heteromeric GIRKs with the G β γ dimer and the interaction forces involved could be different. Knowledge of such differences may clarify the different gating behaviours previously seen with channels of varying composition.

Although the intimate interaction of the GIRK channels with G β γ dimers forms the basis of the channels' gating activity, direct channel interactions with G α also fine-tunes the gating mechanism⁷. How G α provides such control is unknown.

More broadly, it may be possible to design specific molecules that could interfere with channel function by targeting the unique interaction interface of its GIRK2 monomers with G β γ . Such drugs would be desirable because they would not affect other G β γ -dependent signalling events. ■

Eitan Reuveny is in the Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel.
e-mail: e.reuveny@weizmann.ac.il

- Hibino, H. *et al.* *Physiol. Rev.* **90**, 291–366 (2010).
- Whorton, M. R. & MacKinnon, R. *Nature* **498**, 190–197 (2013).
- Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E. J. & Clapham, D. E. *Nature* **325**, 321–326 (1987).
- Wickman, K. D. *et al.* *Nature* **368**, 255–257 (1994).
- Reuveny, E. *et al.* *Nature* **370**, 143–146 (1994).
- Lüscher, C. & Slesinger, P. A. *Nature Rev. Neurosci.* **11**, 301–315 (2010).
- Rubinstein, M. *et al.* *J. Physiol. (Lond.)* **587**, 3473–3491 (2009).

Locomotion dynamics of hunting in wild cheetahs

A. M. Wilson¹, J. C. Lowe¹, K. Roskilly¹, P. E. Hudson^{1†}, K. A. Golabek^{2†} & J. W. McNutt²

Although the cheetah is recognised as the fastest land animal, little is known about other aspects of its notable athleticism, particularly when hunting in the wild. Here we describe and use a new tracking collar of our own design, containing a combination of Global Positioning System (GPS) and inertial measurement units, to capture the locomotor dynamics and outcome of 367 predominantly hunting runs of five wild cheetahs in Botswana. A remarkable top speed of 25.9 m s^{-1} (58 m.p.h. or 93 km h^{-1}) was recorded, but most cheetah hunts involved only moderate speeds. We recorded some of the highest measured values for lateral and forward acceleration, deceleration and body-mass-specific power for any terrestrial mammal. To our knowledge, this is the first detailed locomotor information on the hunting dynamics of a large cursorial predator in its natural habitat.

Measurements of instantaneous speed, acceleration and manoeuvring during athletic competition or hunting are rare^{1–4}, even for humans, horses and dogs, the most studied species. The cheetah (*Acinonyx jubatus*) is acknowledged as the ultimate cursorial predator, and its published⁵ top speed of 29 m s^{-1} is considerably faster than racing speeds for greyhounds² (18 m s^{-1}), horses¹ (19 m s^{-1}) or humans (12 m s^{-1} ; see ‘Analysis of Bolt’s 100m’ at <http://berlin.iaaf.org/records/biomechanics/index.html>). Quantitative measurements of cheetah locomotion mechanics have only been made on captive animals chasing a lure in a straight line, with few studies eliciting speeds faster than racing greyhounds^{6,7}. For wild cheetahs, estimates of speed and track have been made from direct observation or film only, and are limited to open habitat^{8,9} and daylight hours.

Tracking collar design

To collect free-ranging locomotion data on wild cheetahs during hunting in their normal environment, we designed and built a tracking collar similar in size and weight to a conventional wildlife collar^{10,11} (Fig. 1a; mass of 340 g), equipped with a GPS module capable of delivering processed position and velocity data, and raw pseudo-range, phase and Doppler data for individual satellite signals at 5 Hz, and an inertial measurement unit (IMU) consisting of triaxial microelectromechanical systems (MEMS) accelerometers, gyroscopes and magnetometers (Methods). The collar was powered by a rechargeable battery charged from solar cells, plus a non-rechargeable auxiliary battery. Data download and configuration upload was via radio. Collar software monitored the accelerometers to create activity summaries and detect the brief hunting events, buffered accelerometer data to capture the start of hunts, and adapted collar operation to battery voltages, time of day and activity. We increased the effective sample rate of the positioning system to 300 Hz, and reduced noise in the kinematic parameters, by fusing data from GPS and the IMU with a loosely coupled extended Kalman smoother (Methods). This was especially important during hunting because GPS accuracy was degraded both during initialization, and under conditions of high acceleration and high jerk¹².

Collection of hunting data

We recorded GPS–IMU data from 367 runs by three female and two male adult cheetahs (100, 66, 61 and 84, 56 runs respectively) over 17 months. A further 530 runs were identified in the activity data because the collar did not trigger on every run owing to the time of day and conservative trigger thresholds. An episode of feeding after a run indicated hunting success, and was identified in the activity data



Figure 1 | Cheetah with collar and anatomical features contributing to performance. **a**, Cheetah with a mark 2 collar is shown. **b**, Gravitational and centripetal accelerations acting on a turning cheetah; g denotes acceleration due to gravity, $v^2 r^{-1}$ denotes centripetal acceleration, and a is the resultant acceleration (effective gravity). **c**, Non-retractable cheetah claws that enhance grip. **d**, Low posture used in deceleration, which prevents pitching and engages hind limb musculature to absorb kinetic energy.

¹Structure & Motion Laboratory, The Royal Veterinary College, University of London, Hatfield AL9 7TA, UK. ²Botswana Predator Conservation Trust, Private Bag 13, Maun, Botswana. [†]Present addresses: Department of Sport and Exercise Sciences, University of Chichester, College Lane, Chichester, West Sussex PO19 6PE, UK (P.E.H.); Botswana Predator Conservation Trust, Private Bag 13, Maun, Botswana, and Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Oxford OX13 5QL, UK (K.A.G.)

by consistent, low-magnitude acceleration on all three axes¹³ and was confirmed on a subset of hunts with field observations (Methods). Run routes were overlaid on Google Earth to identify terrain. The total number of GPS fixes recorded depended on activity, with an average of 180 ± 171 (mean \pm s.d.) per cheetah per day, and a range of 7 to 1,571.

Runs started with a period of acceleration, either from stationary or slow movement (presumably stalking) up to high speed (Fig. 2). The cheetahs then decelerated and manoeuvred before prey capture. About one-third of runs involved more than one period of sustained acceleration (all 369 runs are presented in Supplementary Video 2). In successful hunts, there was often a burst of accelerometer data after the speed returned to zero, interpreted as the cheetah subduing the prey.

As well as hunting runs, cheetahs play and run from larger predators, but we had insufficient data validated by direct observations to provide secure separation of these activities, although only a few runs did not involve the tight turns and rapid speed changes characteristic of hunting (for example, runs 5, 32 and 49 in Supplementary Video 2). We therefore compared successful hunts to all other runs recorded by the collar. In total, 94 of the 367 runs (26%) were successful hunts.

Including the 530 additional runs detected solely from IMU data did not change the success rate (223 out of 897; 25% success), which is lower than previously reported for individual cheetah^{9,14,15}, perhaps due, in part, to the inclusion of non-hunting runs. Cheetah are reported to move in predominantly open habitats using vegetation-edge to stalk their prey, often at dawn and dusk^{8,9,14–16}. Although almost half of the runs here occurred at/after dawn, runs occurred throughout the day and night (Fig. 3e). The individual cheetahs varied in their predilection for running in open grassland or dense shrub (Supplementary Fig. 6). On average, the cheetahs ran most often in open habitat (48%, 176 of 367 runs); 28% of runs occurred in open shrub/around large trees, and 24% occurred within dense vegetation. Only 20% of runs occurring in the open grasslands were identified as successful hunts, compared with 31% of runs in dense cover. This difference in outcome was not significant ($P = 0.054$, chi-squared test) and is confounded by individual variation and habitat, but it does demonstrate that cheetahs do hunt successfully in all terrains^{8,15}. Vegetation may confer an advantage by permitting stalking and limiting prey options for escape by manoeuvring; however, there was little difference in the distance or speed between terrains (Supplementary Table 1).

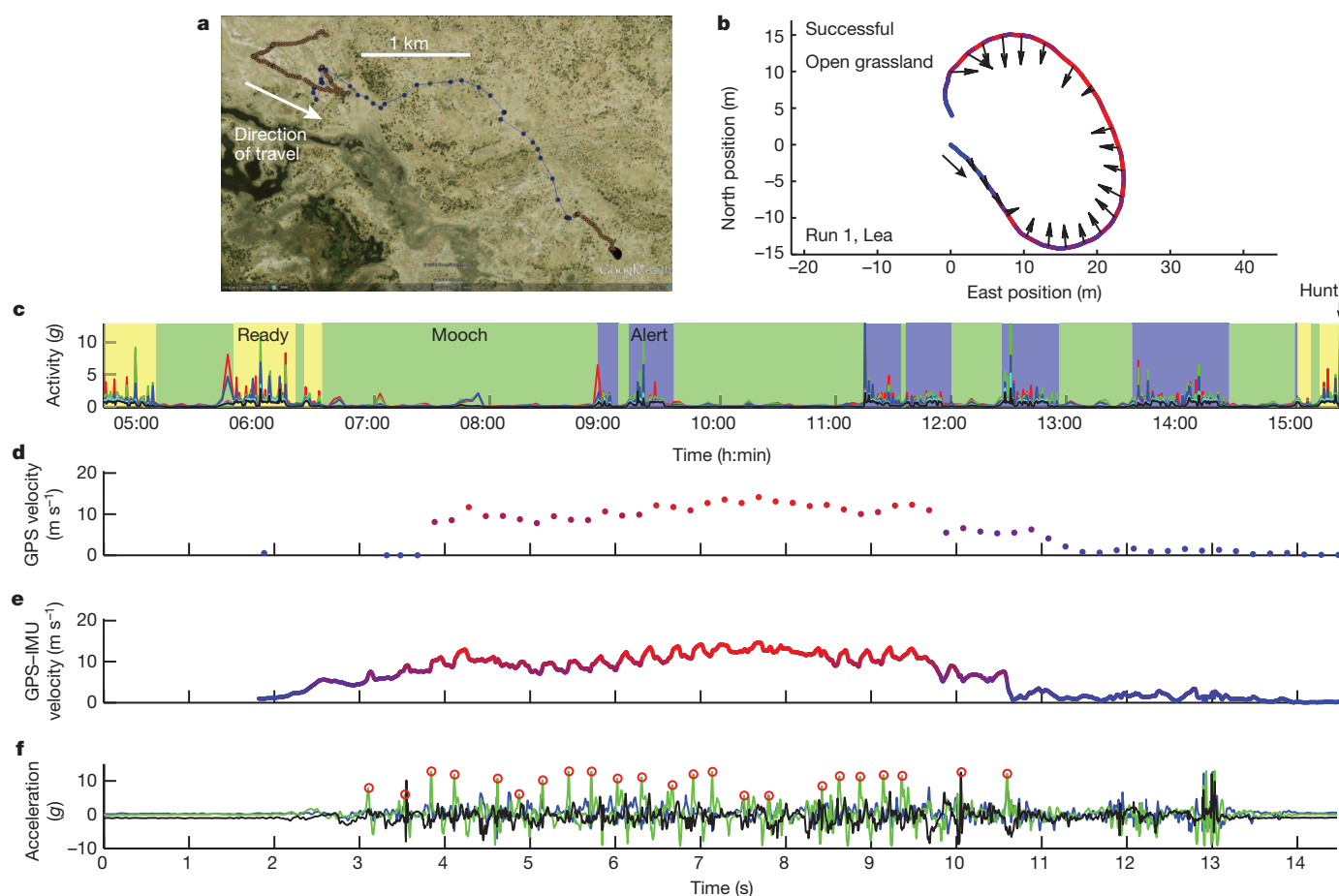


Figure 2 | An example day and hunt. **a**, Track of cheetah over 11 h (GPS data are available as a Google Earth file in Supplementary File 1). Each circular mark represents a GPS-derived position. Cheetah track and marks are colour-coded to collar state (detailed in Supplementary Fig. 1) as follows: alert, blue; mooch, green; ready, yellow; chase, red. **b**, Hunt track magnified from bottom right of **a**, hunt track is anticlockwise and marked with an arrow. Warmer (bright red) colours on track represent higher speed. **c**, Activity summary calculated in the collar from the accelerometer (Methods) for the 11-h period shown in **a**; shaded regions of the graph represent collar states as labelled. Line colours: peak accelerometer signal amplitude recorded in each 30-s period X, blue; Y, green; Z, red; mean of peak amplitude values extracted for each 2-s in each 30-s (that is, 15 bins) period X, cyan; Y, magenta; Z, black. The relative values for each axis differentiate

between a single high-acceleration cycle and consistent movement in the 30-s window. Coordinate system: X lateral, positive left, Y fore-aft, positive forwards, Z vertical, positive upwards. Time is local (coordinated universal time (UTC) + 2 h). 'Hunt' time is labelled. **d**, Doppler-derived velocity profile for hunt determined by the GPS receiver at five updates per second. **e**, GPS-IMU-derived velocity profile for the chase; in **b**, **d** and **e** warmer (bright red) colours represent faster speeds. **f**, Accelerometer data recorded at 300 Hz for chase; X, blue; Y, green; Z, black. Red circles indicate forward acceleration peak used as event marker for stride cutting at, approximately, hindlimb foot contact. The high accelerations at zero velocity at $t = 12–13$ s suggest subduing prey and a successful hunt. An animation of a hunt is in Supplementary Video 1, plots of further runs are available in Supplementary Fig. 5, and all runs are in Supplementary Video 2.

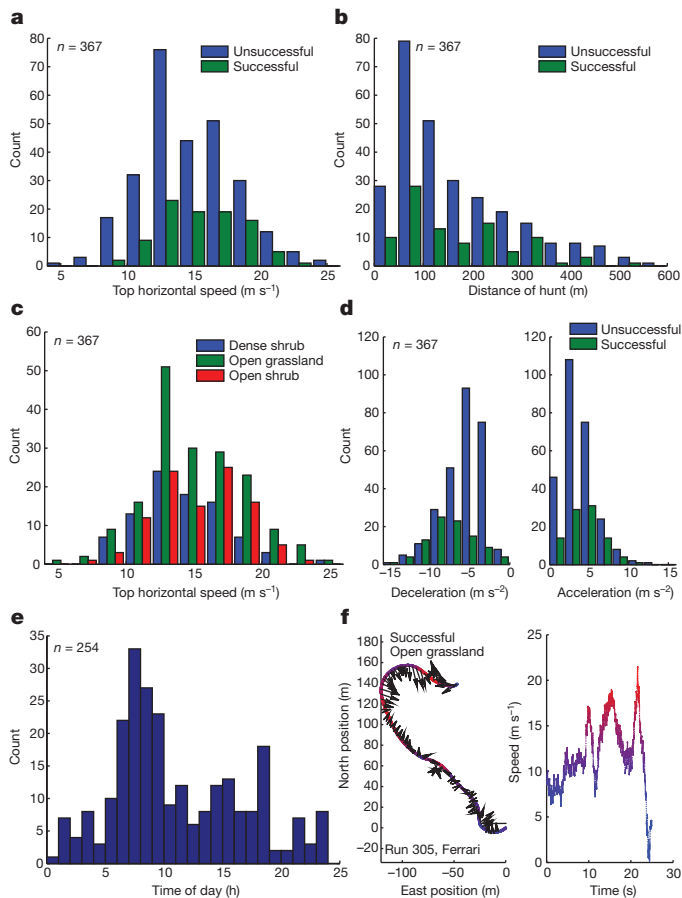


Figure 3 | Descriptive hunt statistics. **a**, Top speed, averaged over a stride, reached in each run colour-coded for outcome. **b**, Distance covered in each run. **c**, Top speed in each run coded for terrain type. **d**, Peak acceleration and deceleration recorded in each run. **e**, Plot of time of day of runs recorded in period when collar was set to trigger at any time of day, time local. **f**, Example hunt file colour-coded for speed (bright red denotes fastest), and with horizontal acceleration vectors drawn, to scale, for each stride. $n = 367$ (**a–d**) and $n = 254$ (**e**).

Description of hunts

The average run distance was 173 m (± 116 m) (Fig. 3b) though recorded run distance will be shorter than the true value in the runs where the start of the run was missed (Methods, Supplementary Video 2). The longest runs recorded by each cheetah ranged from 407 to 559 m; the mean run frequency (including information from activity data) was 1.3 times per day, so, even if some hunts were missed, high speed locomotion only accounted for a small fraction of the 6,040-m average daily total distance covered by the cheetahs. The mean top speed was 14.9 ± 3.4 m s⁻¹ and was usually only sustained for 1–2 s. The highest speed we recorded was a stride-averaged 25.9 m s⁻¹ in run 250 (Fig. 3a, c and Supplementary Video 2). The top speeds attained by the other cheetahs were 25.4, 22.0, 21.1 and 20.1 m s⁻¹. The cheetahs studied here mostly hunted impala (*Aepyceros melampus*)¹⁷, which made up 75% of their diet, although one male cheetah (Qamar), which frequently hunted in thicker vegetation (Supplementary Fig. 6), never exceeded 20.1 m s⁻¹ and was often observed on warthog (*Phacochoerus africanus*) kills. Cheetah hunting the (anecdotally) faster Thompson's gazelle (*Eudorcas thomsonii*) on open East African savannah may use higher speeds.

Successful hunts involved greater deceleration on average (-7.5 m s⁻² versus -5.5 m s⁻²; $P < 0.05$; Fig. 3d), but there was no significant difference in peak acceleration (Fig. 3d), distance travelled (Fig. 3b) number of turns (6.7 versus 6.5) or total turn angle (347° versus 260°) (generalized linear mixed model (GLMM); Methods). This indicates

that outcome was determined in the final stages of a hunt rather than hunts being abandoned early to save energy or reduce risk of injury, and the higher deceleration values may reflect actual prey capture. Equivalent locomotion and outcome data for coalition-hunting cheetah might clarify the importance of the final manoeuvring phase in hunt outcome.

Comparison with other athletic animals

The greatest acceleration and deceleration values were almost double values published for polo horses¹⁸ and exceeded the accelerations reported for greyhounds at the start of a race¹⁸. The cheetahs sped up by up to 3 m s⁻¹ and slowed by up to 4 m s⁻¹ in a single stride (Supplementary Fig. 5d). Mass-specific change in kinetic energy over a stride (Fig. 4c and Supplementary Fig. 7) exceeded 30 J kg⁻¹ stride⁻¹ across the broad speed range of 10 to 18 m s⁻¹. On the basis of forward acceleration, the greatest stride-averaged whole animal powers often exceeded 100 W kg⁻¹ (body mass) (Fig. 4d), and also occurred between 10 and 18 m s⁻¹. For comparison, we calculated a stride-averaged power of 25 W kg⁻¹ for Usain Bolt's 9.58-s 100-m world record (Methods and <http://berlin.iaaf.org/records/biomechanics/index.html>), consistent with other measurements on human sprinters¹⁹; polo horses achieve 30 W kg⁻¹ (ref. 18) and racing greyhounds 60 W kg⁻¹ (ref. 18).

The locomotor (limb and back) muscle accounts for $45 \pm 4\%$ of body mass^{20,21} in captive cheetah. The wild cheetahs had similar limb and back lengths to those captive cheetahs, but were heavier at 53 kg versus 33 kg (means, $n = 5, 5$), and visibly more muscled (mean mid-thigh girth 540 mm versus 450 mm, $n = 5, 5$), so much of their body mass is locomotor muscle. Major propulsive muscles such as the hamstrings (biceps femoris, semimembranosus and semitendinosus) at the hip and gastrocnemius at the tarsus have 64% and 60% longer moment arms, respectively, than in the greyhound and similar muscle fibre lengths²¹. Stride frequency and posture are similar at the same speed in the two species⁷ so the muscle sarcomeres (and fibres) will be shortening considerably faster in the cheetah than in the greyhound at the same speed (like the engine of a car in a lower gear). This fast muscle contraction velocity will enable large muscle powers and hence deliver the very large acceleration powers observed²². The high muscle speed and power are consistent with our measurements on contracting skinned fibres from cheetahs²³. The cheetah deceleration magnitudes (Figs 3d and 4b), cycle works (Fig. 4c) and powers (Fig. 4d) were greater than during acceleration and up to three times higher than polo horses¹⁸; however, comparative figures are sparse. Cheetah can crouch to engage locomotor muscle to enable these deceleration magnitudes (Fig. 1d), and sliding or colliding with the prey may dissipate some energy.

Grip and manoeuvrability key to hunting success

Hunts involved considerable manoeuvring, with maximum lateral (centripetal) accelerations often exceeding 13 m s⁻² at speeds less than 17 m s⁻¹ (Fig. 4e, f; polo horses achieve 6 m s⁻²; ref. 3). A lateral acceleration of 13 m s⁻² (Fig. 1b) requires a coefficient of friction with the ground of at least 1.3. Ridged footpads and substantial claws²⁴ (Fig. 1c) act as cleats to augment friction and deliver this level of grip. The maximum centripetal acceleration observed was smaller at speeds greater than 17 m s⁻¹ (Fig. 4e), which may be behavioural in origin; that is, cheetahs do not perform tight turns at their highest speeds. Studies on other animals show that, although grip limits turning performance at low and moderate speed, a model based on the capacity of the limbs to withstand the combination of centripetal acceleration and gravity (Fig. 1b) is appropriate to account for reduced speed on bends in humans, mice and racehorses^{3,25–27} but not greyhounds². The dashed line labelled LFL (leg force limit) in Fig. 4e is calculated using published models^{3,25,27}, published stride data⁷ and the maximum speed recorded here. The equations and assumptions are presented in the Supplementary Information. The LFL line seems to follow the upper bound of the data points at higher speeds, but confident verification would,

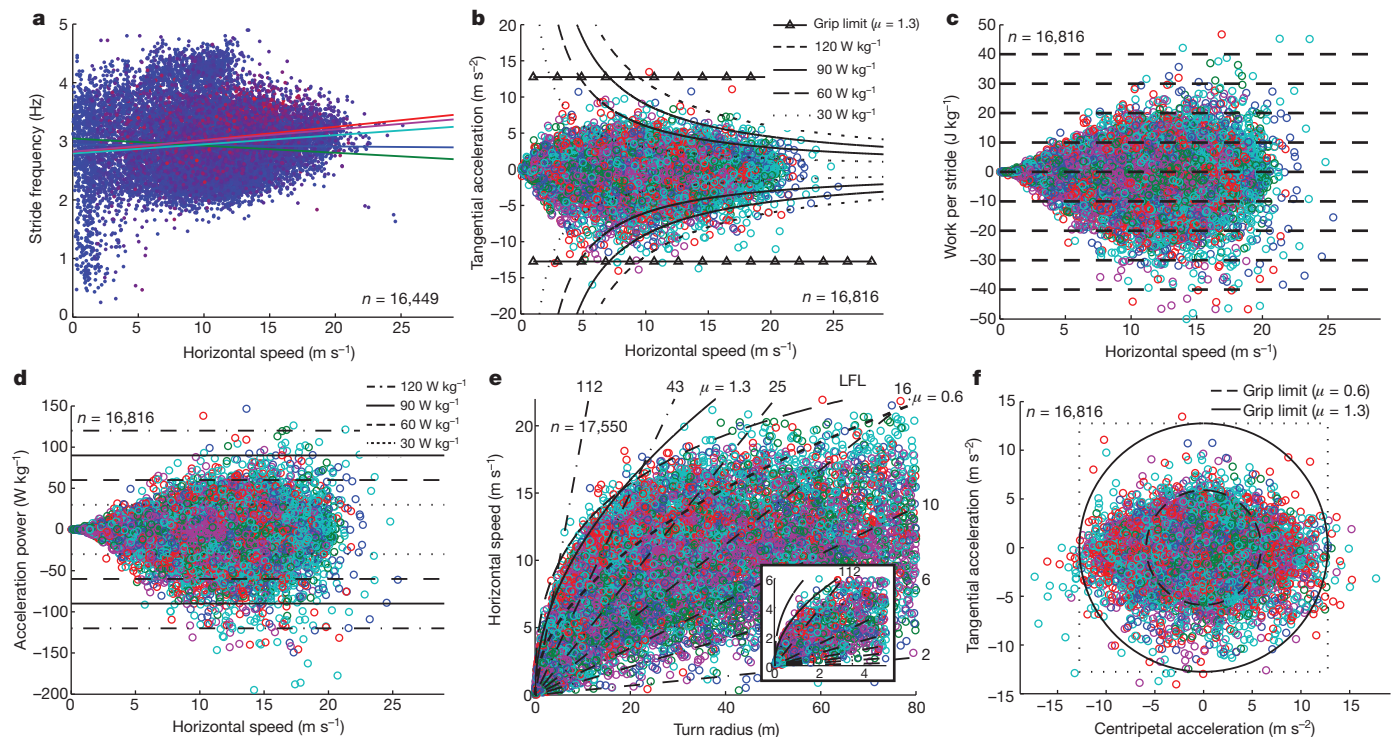


Figure 4 | Performance summary. **a**, Stride frequency plotted against speed; each point is colour-coded for tangential (forwards) acceleration, bright red points represent the greatest forward acceleration, and are plotted last (on top). Lines are linear regression of stride frequency against speed for each individual cheetah. **b**, Tangential (forwards, positive) acceleration and deceleration (y axis) against speed (x axis). Horizontal lines represent acceleration and deceleration of 13 m s⁻², equating to the proposed grip limit of 1.3 (see text). Curved lines represent stride-averaged whole-body powers of ± 30 , 60, 90 and 120 W kg⁻¹; points outside the outer dashed line equate to a mean stride power in excess of ± 120 W kg⁻¹. **c**, Body, mass specific, horizontal kinetic energy change performed in each stride (work per stride). **d**, Stride-averaged whole-body acceleration power plotted against speed, with horizontal lines showing powers of ± 30 , 60, 90 and 120 W kg⁻¹. **e**, Horizontal speed against turn radius, region around origin magnified in inset. Slanting straight lines show different rates of heading change in degrees per second, with values (2, 6, 10, 16, 25, 43 and 112) at

the top of the line. The solid curved line ($\mu = 1.3$) represents a grip limit/coefficient of friction of 1.3; the curved shorter-dashed line ($\mu = 0.6$) denotes the 0.6 grip limit reported for polo horses³; points above each line require a higher grip level. The curved longer-dashed line (LFL) represents a limit to turning defined by the maximum force the legs can withstand. **f**, Plot of tangential acceleration against lateral acceleration. Total horizontal acceleration is the distance from the origin, circles represent mean total horizontal acceleration of 6 and 13 m s⁻² (equating to average grip limits of 0.6 and 1.3). Each point on each plot represents data centred on a single stride, with data smoothed over three strides. Points are colour-coded by individual, except in plot **a**. The number of strides from each cheetah were 5,031, 4,022, 3,211, 2,657 and 1,895 giving a total n of 16,816 for plots **b**, **c**, **d** and **f**. The total n is given in each plot and was slightly different for plots **a** and **e** owing to the mathematics of generating those plots but the individual contributions were in proportion.

however, require stance times or limb forces during manoeuvring²⁶. When combined with gravity, a lateral acceleration of 13 m s⁻² equates to a 66% increase in the cheetah's effective weight and hence average limb force (Fig. 1b). Cheetahs have relatively large limb bone cross-sectional areas (compared with greyhounds^{20,21}), which may be an adaptation to resist the large peak limb forces that occur during high speed manoeuvring.

The cheetah should run little faster than its prey in the manoeuvring phase of the hunt^{28,29} if it is to capture an agile and quick-turning prey. A cheetah running at 25.9 m s⁻¹ with the maximal observed lateral acceleration of 13 m s⁻² would have a turn radius of 52 m and would take 6 s to perform a 180° turn ($\pi r v^{-1}$)—peak running speed is therefore unlikely to be, and was not found to be, a feature of the final stage of successful hunts. A cheetah can slow by 4 m s⁻¹ in a stride (Supplementary Fig. 5d), and the cheetahs often decelerated sharply before turning, which would enable much tighter turns. Slowing from 16 m s⁻¹ to 4 m s⁻¹ (three strides, 1 s) would drop the turn radius with $v^2 r^{-1} = 13$ (lateral acceleration of 13 m s⁻²) from 19.7 m to 1.2 m, and heading velocity ($v r^{-1}$) would rise from 46 to 190° s⁻¹. This demonstrates the value of slowing down before manoeuvring. The cheetahs did not use highest tangential and centripetal accelerations simultaneously, consistent with grip limiting maximal horizontal acceleration (there are few data points in the corners of the square in Fig. 4f). Rapid deceleration would unload the hindquarters, which could result in yaw

instability when manoeuvring because the centre of mass (COM) is behind the forelimbs (like a ground loop in a tail wheel aircraft). The pitch limit proposed in ref. 18 may apply at low speed, but insufficient low-speed data exist to consider this further, and it can be circumvented by posture due to the cheetah's flexible spine (Fig. 1d). The active movements of the high-inertia tail that are observed in wildlife documentaries will help in positioning and banking the body (and limbs) to apply appropriate forces to prevent this and for turn initiation and manoeuvring.

Perspective

Equivalent data for other wild cursorial species would enhance what we know about natural speed, agility, endurance and locomotor physiology, and provide detailed information on ranging behaviour in the wild. For example, such fine-scale data on habitat selection by endangered species detailing where animals are commuting, hunting and resting will be informative when attempting to evaluate landscape scale connectivity, corridors and wildlife-protected areas. Tightly coupled GPS-IMU processing can deliver 0.2-m position accuracy (the level of individual shrubs and footfalls) during hunts, enabling detailed analysis of context variables (such as habitat characteristics and prey visibility), modes of hunting success and failure, and the effect of slope, camber and foot-surface interaction on stride-by-stride performance. These data on hunt environment would inform about

the determinants of preferred hunting habitats, risk of injury (of paramount importance for solitary predators), risk of detection by kleptoparasites (open versus closed habitat), available palatable grazing and habitat-dependent risk of predation (detection).

METHODS SUMMARY

Collars moved between six operating states depending on the time of day, the activity level of the cheetah, and battery voltages (Supplementary Fig. 1). If the cheetah were active (detected via accelerometers) at a time when hunting was likely, accelerometer data samples were continuously buffered in memory, and the GPS module was regularly triggered ('refreshed') to maintain an internal state ready for immediate start-up. When a run started, GPS data at 5 Hz and full IMU data at 300 Hz were recorded. The GPS-IMU data were post-processed in a loosely coupled extended Kalman smoother optimized for sensor characteristics (Methods) and cheetah dynamics. Horizontal position error (median stride-wise standard deviation (s.d.), $n = 45,851$) was reduced from 5.05 m (pure GPS data) to 0.67 m in the smoothed solution. Speed error was reduced from 1.23 m s^{-1} to 0.34 m s^{-1} (Supplementary Fig. 3). The initial seconds of the run were reconstructed by open-loop inertial integration, backwards in time, using buffered IMU data and smoothed GPS-IMU data for initial conditions. Data were segmented into strides using the horizontal acceleration signal, and a rolling average was applied to the stride duration, speed and heading rate data (methods) to ensure that cutting did not result in erroneous extreme values in these or derived parameters (Supplementary Fig. 4). Activity summaries, based on accelerometer readings, were recorded for each 30-s period throughout the rest of the day, with a GPS position every 5 min when the cheetah was on the move. The dynamic performance of the collar for track and speed was verified by running a dog on a beach (Supplementary Fig. 2); footprint position in the sand was determined using survey-grade GPS, and footfall time from GPS time-stamped high-speed video.

Full Methods and any associated references are available in the online version of the paper.

Received 15 June 2012; accepted 17 May 2013.

- Spence, A. J., Thurman, A. S., Maher, M. J. & Wilson, A. M. Speed, pacing strategy and aerodynamic drafting in Thoroughbred horse racing. *Biol. Lett.* **8**, 678–681 (2012).
- Usherwood, J. R. & Wilson, A. M. Biomechanics: no force limit on greyhound sprint speed. *Nature* **438**, 753–754 (2005).
- Tan, H. & Wilson, A. M. Grip and force limits to turning performance in competition horses. *Proc. R. Soc. Lond. B* **278**, 2105–2111 (2011).
- Self, Z. T., Spence, A. J. & Wilson, A. M. Speed and incline during Thoroughbred horse racing: racehorse speed supports a metabolic power constraint to incline running but not to decline running. *J. Appl. Physiol.* **113**, 602–607 (2012).
- Sharp, N. C. C. Timed running speed of the cheetah (*Acinonyx jubatus*). *J. Zool.* **241**, 493–494 (1997).
- Hildebrand, M. Further studies on locomotion of the cheetah. *J. Mamm.* **42**, 84–91 (1961).
- Hudson, P. E., Corr, S. A. & Wilson, A. M. High speed galloping in the cheetah (*Acinonyx jubatus*) and the racing greyhound (*Canis familiaris*): Spatio-temporal and kinetic characteristics. *J. Exp. Biol.* **215**, 2425–2434 (2012).
- Bissett, C. & Bernard, R. T. F. Habitat selection and feeding ecology of the cheetah (*Acinonyx jubatus*) in thicket vegetation: is the cheetah a savanna specialist? *J. Zool.* **271**, 310–317 (2006).
- Eaton, R. L. *The Cheetah: The Biology, Ecology, and Behaviour of an Endangered Species* Ch.3 (Van Nostrand Reinhold, 1973).
- Cagnacci, F., Boitani, L., Powell, R. A. & Boyce, M. S. Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Phil. Trans. R. Soc. B* **365**, 2157–2162 (2010).
- Tomkiewicz, S. M., Fuller, M. R., Kie, J. G. & Bates, K. K. Global positioning system and associated technologies in animal behaviour and ecological research. *Phil. Trans. R. Soc. B* **365**, 2163–2176 (2010).
- Tan, H., Wilson, A. M. & Lowe, J. Measurement of stride parameters using a wearable GPS and inertial measurement unit. *J. Biomech.* **41**, 1398–1406 (2008).
- Grünwälder, S. *et al.* Movement activity based classification of animal behaviour with an application to data from cheetah (*Acinonyx jubatus*). *PLoS ONE* **7**, e49120 (2012).
- Caro, T. M. *Cheetahs of the Serengeti Plains* (Univ. Chicago Press, 1994).
- Mills, M. G. L., Broomhall, L. S. & du Toit, J. T. Cheetah *Acinonyx jubatus* feeding ecology in the Kruger National Park and a comparison across African savanna habitats: is the cheetah only a successful hunter on open grassland plains? *Wildl. Biol.* **10**, 177–186 (2004).
- Durant, S. M. Competition refuges and coexistence: an example from Serengeti carnivores. *J. Anim. Ecol.* **67**, 370–386 (1998).
- Broekhuis, F., Cozzi, G., Valeix, M., McNutt, J. W. & Macdonald, D. W. Risk avoidance in sympatric large carnivores: reactive or predictive? *J. Anim. Ecol.* <http://dx.doi.org/10.1111/1365-2656.12077> (20 May 2013).
- Williams, S. B., Tan, H., Usherwood, J. R. & Wilson, A. M. Pitch then power: limitations to acceleration in quadrupeds. *Biol. Lett.* **5**, 610–613 (2009).
- Cavagna, G. A., Komarek, L. & Mazzoleni, S. The mechanics of sprint running. *J. Physiol.* **217**, 709–721 (1971).
- Hudson, P. E. *et al.* Functional anatomy of the cheetah (*Acinonyx jubatus*) hindlimb. *J. Anat.* **218**, 363–374 (2011).
- Hudson, P. E. *et al.* Functional anatomy of the cheetah (*Acinonyx jubatus*) forelimb. *J. Anat.* **218**, 375–385 (2011).
- Curtin, N. A., Woledge, R. C. & Aerts, P. Muscle directly meets the vast power demands in agile lizards. *Proc. R. Soc. Lond. B* **272**, 581–584 (2005).
- West, T. G. *et al.* Power output of skinned skeletal muscle fibres from the cheetah (*Acinonyx jubatus*). *J. Exp. Biol.* <http://dx.doi.org/10.1242/jeb.083667> (11 April 2013).
- Russell, A. P. & Bryant, H. N. Claw retraction and protraction in the carnivore: the cheetah (*Acinonyx jubatus*) an atypical felid. *J. Zool.* **254**, 67–76 (2001).
- Usherwood, J. R. & Wilson, A. M. Accounting for elite indoor 200 m sprint results. *Biol. Lett.* **2**, 47–50 (2006).
- Chang, Y. H. & Kram, R. Limitations to maximum running speed on flat curves. *J. Exp. Biol.* **210**, 971–982 (2007).
- Greene, P. R. Running on flat turns: experiments, theory, and applications. *J. Biomech. Eng.* **107**, 96–103 (1985).
- Schaller, G. B. Hunting behaviour of the cheetah in the Serengeti National Park, Tanzania. *Afr. J. Ecol.* **6**, 95–100 (1968).
- Howland, H. C. Optimal strategies for predator avoidance: the relative importance of speed and manoeuvrability. *J. Theor. Biol.* **47**, 333–350 (1974).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Amos and M. Dickson for fabricating collars, and F. Broekhuis, R. Furrer and N. Jordan for working with us in the study area; P. Apps for many discussions; T. Hubel and A. Wills for helping to collect and analyse validation data; P. Apps, J. Usherwood and A. Wilson for comments on the manuscript; and the EPSRC (EP/H013016/1), BBSRC (BB/J018007/1) and DARPA M3 Program (W91CRB-11-C-0048, with Boston Dynamics) for funding. This work was approved by RVC Ethics & Welfare Committee and was carried out under a Botswana Government Research Permit held by J.W.M. and Botswana Veterinary Registration held by A.M.W.

Author Contributions A.M.W. conceived, designed and led the study, A.M.W. and J.C.L. designed the collars, J.C.L. developed collar hardware and software, K.R. developed GPS-IMU filtering, K.R., P.E.H. and A.M.W. analysed data, A.M.W. wrote the paper with input from all authors, K.A.G. and J.W.M. organized field work, monitored animals and downloaded data. A.M.W. performed veterinary work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.W. (awilson@rvc.ac.uk).

METHODS

Animals. The cheetahs used in this study were part of a continuing study by Botswana Predator Conservation Trust (<http://www.bpctrust.org>) in the Okavango Delta region of Northern Botswana. Initially, three 'mark 1' prototype collars were fitted to three cheetahs in July 2011. All collars successfully collected data as intended, two collars for 7–9 months whereas the third suffered a memory card failure after 6 months. Three collars of a new 'mark 2' design were used in April 2012, and two more collars in July 2012 (fitted to the original three cheetahs plus two new individuals). Data were again successfully collected from these collars, and they continue in operation.

The cheetahs were immobilized by free darting from a vehicle by A.M.W. using medetomidine (2 mg) and ketamine (80–120 mg) and reversed after 60 min with 10 mg atipamezole. While sedated, dimensions including limb lengths, thigh girths and back lengths and body mass were recorded. Collar data were downloaded by radio link every few weeks to a ground vehicle or a light aircraft.

Collar design and fabrication. The major design challenges included the measurement and logging of data at a sufficiently high rate and accuracy, timely remote retrieval of substantial volumes of data from the collar and maintaining the very low average power consumption required in a wildlife collar. To conserve power, careful management of the internal readiness of the GPS subsystem allowed this and other sensor systems to be started quickly enough to capture data at maximum rate only during these events.

All collars were constructed in-house. In the original collars (mark 1, used in 2011), a commercial radio-tracking collar (Sirtrack, New Zealand) was used as a base, our custom electronics package being mounted on the top of the collar in a clear cast resin case and wired to the collar's original battery box at the bottom of the collar. The revised mark 2 collars (Fig. 1a) were entirely constructed in-house, with a revised lower-profile electronics enclosure (cast from polyurethane resin using a silicon mould and a rapid prototyped former; Aprocas GmbH) and a vacuum-formed polycarbonate battery box holding larger rechargeable and back-up battery in potting compound. The actual electronics package was similar on both versions, with an identical chip set as described below, and with almost identical software functionality. Collar mass was approximately 340 g.

Collar design: electronics payload. The collar circuit was based around a low-power MSP430 16-bit microcontroller (Texas Instruments), running custom software written in the 'C' programming language developed using an integrated development system from IAR Systems. The microcontroller contains several internal peripheral blocks, including an 8-channel 12-bit analogue-to-digital converter (ADC), four serial communications modules, plus various timers, general-purpose digital input and output lines, and other support modules. A connected 2-GB micro-SD flash memory card (Sandisk) provided data storage.

GPS position was obtained from an LEA-6T GPS module (u-Blox AG). In addition to internally computed position and velocity, the module is able to generate raw pseudo-range, phase and Doppler data for the signal from each satellite enabling detailed GPS performance evaluation, and use of customized differential techniques for increased accuracy. The data rate was five position, velocity and raw data points per second during continuous operation (for example, during a chase).

The collar circuit also included an inertial measurement suite, based on MEMS devices. Acceleration was measured using an MMA7331 three-axis accelerometer module (Freescale Semiconductors), providing acceleration with a ± 12 g range. The roll and pitch rotation rate was measured by a dual-axis gyroscope (ST Microelectronics), and yaw rotation rate by a single-axis gyroscope (ST Microelectronics), both set to the $2,000^\circ \text{ s}^{-1}$ range. Sensor outputs were filtered by simple single-pole analogue filters (100 Hz knee), and then sampled by the microcontroller ADC at 300 or 100 samples per second (Accelerometers or Gyroscopes, respectively). Three-hundred hertz was chosen as giving an overhead to a frequency of 30 Hz; that is, $1/\text{minimum published stance time}^2$. A three-axis magnetometer (Honeywell), connected via I²C, provided magnetic compass functionality at 12 measurements per second.

Primary communication with the collar, for tasks such as data file download and configuration file upload, was via a 2.4-GHz chirp-spread-spectrum communication module (Nanotron Technologies GmbH), communicating at 1 Mbit per second using a custom communications protocol. A 173-MHz VHF radio transmitter (Radiometrix) provided longer-range transmission of current GPS-derived position, for tracking purposes. An original equipment manufacturer (OEM) conventional wildlife tracking transmitter in the 149-MHz band (Sirtrack) facilitated long-range animal location using conventional direction-finding techniques.

Collar design: power. Primary power supply for the collar was a 900 mAh lithium-polymer rechargeable battery (Active Robots), charged by a solar cell array consisting of 10 monocrystalline silicon solar cells (Ixys Koria). On the mark 2 collars, a 13 Ah lithium thionyl chloride primary battery (Saft) provided

a back-up power source (on the original collars, a 7.7 Ah lithium thionyl chloride primary battery was used). Both battery voltages, together with the charge current from the solar cell array, were measured by the microcontroller, which switched the collar electrical load from one battery to the other depending on battery state.

Collar design: software states and movement detection. In operation, the collar software moved between several different operating 'states', the particular state at any moment being dependent on a combination of animal activity level (measured using the accelerometers) and time of day (from a GPS-synchronised software clock). Each state required a different mix of hardware sub-systems to be powered on or off, and different intervals between GPS module operation, and thus the power consumption of the collar varied depending on the operating state. Thus, the inevitable compromise between average power consumption on the one hand, and quantity and resolution of data gathered on the other, could be optimized by setting the parameters for the state transitions. The different operating states and associated average power consumption for the collar are summarized in Supplementary Fig. 1.

To keep the average power consumption as low as possible, the collar would generally default to operating in state 1 ('alert' state). In this state, to detect when the cheetah was moving, the accelerometer was sampled at 30 Hz for a period of 10 s in every minute. Within each 10-s sampling period, the peak-to-peak acceleration was computed for each axis every 2 s, and an accumulator incremented by a specified value for each 2-s window in which the peak-to-peak acceleration exceeded a pre-set threshold; For each 2-s window in which the peak-to-peak acceleration did not exceed the threshold the accumulator was decremented by a (different) specified value. Thus, periods of movement could be given higher 'weight' than periods of no movement or vice versa to identify stalking. If the accumulator total exceeded a specified value, the cheetah was deemed to be consistently moving and the collar switched to a higher operating state, the exact state depending on time of day. A similar algorithm with different weights and thresholds was then used to determine when the animal had settled back to rest, at which time a switch back to the lower state was executed.

When consistently moving between local times of 06:00 and 09:00, and 17:00 and 19:00 (times when hunting was most likely from previous work), the operating state would transition to state 3 ('ready' state). The GPS was refreshed every 30 s and position recorded every 60 s. Accelerometer data were recorded into a circular buffer at 100 Hz, the buffer storing the latest 3 s of data. If the fore-aft accelerometer data then exceeded a threshold equivalent to galloping, state 4 ('chase' state) would be entered. The buffered data were stored and 5 Hz GPS data, 300 Hz accelerometer, 100 Hz gyroscope and 12 Hz magnetometer data recorded. A record was defined as valid if five further peaks (strides) were detected, and then recording would continue until there were no peaks above the threshold for 5 s. When moving consistently but outside of the peak hunting times, the lower-powered state 2 ('mooch' state) would be invoked, with GPS positions being taken every 5 min and simple activity measurements being taken as described below. The GPS delivered a first fix in 1.30 s after triggering (median), accurate position data (<10 m s.d.) after 1.58 s, and full rate data (5 Hz) after 5.4 s (Supplementary Fig. 3). The unexpectedly long delay in the GPS module delivering 5 Hz data prevented open-loop GPS-IMU integration back to the beginning of the run in some cases. This is why many runs in Supplementary Video 2 do not start at low speed.

Collar power handling and power consumption. Average collar power consumption varied between individual animals (owing to differing patterns of activity and hence a different distribution of collar operating states), but was typically around 4 mA when averaged over 24 h. The main contributor to this average was the time spent in the ready state when the animal was active during hunting times of day (Supplementary Fig. 1), in which average consumption was around 16 mA with a 30-s GPS refresh time. By comparison, the time spent in the mooch state (animal active but outside hunting times) had a lower consumption of about 5 mA, whereas 'sleep' or alert states (animal inactive) contributed only about 0.6 mA. The 'chase' state, used only when the animal is running, required some 90 mA, but time spent in this state was very small. Solar charge currents ranged from 35 mA with the animal in full sunlight, to typically 10 mA in dappled shade and almost zero in deeper shade. Average charge current over a 24-h period was typically 2 mA, with some variation between animals due to terrain preferences, indicating little time spent in full sunlight even in the winter study period. The solar cells, via the rechargeable battery, contributed roughly 75% of the collar power, the remainder being supplied by the non-rechargeable battery. Collar battery life was predicted at approximately one year with these settings, but was very dependent on collar settings and animal behaviour.

On cheetahs four and five, the ready state GPS refresh interval was changed from 30 s to 300 s—this resulted in a typical power saving of around 30% over a 24-h period, with unexpectedly little effect on GPS start-up time (Supplementary Fig. 3f). We reduced power consumption on mark 2 collars (254 runs) by not

pre-buffering data, and moving directly from mooch to chase state (and allowing this to happen at any time of day, enabling Fig. 3e to be generated), so that IMU data logging began on the first accelerating stride when the cheetah was already in motion. The time that could be recovered through backwards integration was therefore reduced, and the first 1–2 acceleration strides lost.

Collar design: generation of activity summaries. Throughout all states, a background measurement of animal activity was also recorded. For every 2-s ‘window’, the maximum peak-to-peak acceleration range is recorded separately for all three accelerometer axes. After 15 ‘windows’ have passed, an activity record is generated, containing GPS time, the largest X, Y and Z peak-to-peak acceleration amplitudes seen in any of the 15 windows, and the average of the 15 2-s peak-to-peak X, Y and Z accelerations amplitudes. This enabled differentiation of transient high acceleration events and consistent activity. This record is generated continuously in the mooch and ready state, every 3 min in the alert state, and every 30 min in the sleep state. Amplitudes are higher than body acceleration, because the collar can move relative to the centre of mass.

All settings that affected the state transitions (times, acceleration thresholds, and so on), and many other settings besides, could be modified by uploading a new configuration file over the 2.4-GHz communications link. In addition, a complete new version of the collar firmware could be uploaded over this link, allowing for in-field program updates while the collar is on the animal.

Sensor fusion and signal processing to capture hunting dynamics. In the collar data collected here, the power management features used gave different sampling rates for accelerometer (300 Hz) and gyro (100 Hz) in the chase state. To capture the full acceleration profile within the microcontroller, 3 s of accelerometer measurements were continually buffered in ready state at a reduced sampling frequency (100 Hz) and recorded when entering the chase state (gyro-power consumption was too high to permit continuous pre-buffering). GPS position and velocity measurements were usually (but not always) available within 1 s after entering the chase state (Supplementary Fig. 3).

The unique characteristics of these data required a custom-designed GPS–INS (inertial navigation system) integration method written in Visual C++ and MATLAB. Calibrated IMU measurements were first linearly interpolated to 300 Hz. Orientation changes were assumed to be minimal during the buffer period, and hence the unmeasured gyro angular rates assumed to be zero. GPS and IMU measurements were fused using a 12-state extended Kalman filter³⁰ in loosely coupled architecture. The total state formulation used propagates position, velocity and orientation states with time using the IMU measurements in a simplified form of the strap-down inertial navigation equations³¹. The associated process noise was estimated from the known error characteristics of the inertial sensors used. GPS position and velocity updates were used as measurement updates, and receiver accuracy data for each fix used to estimate measurement noise to appropriately weight the GPS to the inertial solution.

The filter was run in reverse time from the last GPS observation of each run to the beginning of the buffered inertial data. During the short time period in which only inertial data was present (throughout buffer and between GPS measurements), the filter propagation was equivalent to open-loop inertial navigation. The filter was initialised using last GPS position and velocity data, and Euler angles assumed zero with covariances appropriate for the uncertainty in that assumption. A Rauch-Tung-Striebel (RTS) smoother³² was then applied in forward time on the Kalman-filtered data. This is equivalent to combining backward and forward solutions, effectively halving the open-loop INS integration period between GPS observations. It was not always possible to reconstruct the period before the first GPS observation, as this period was often too long or the accuracy of the initial GPS observations insufficient (Supplementary Fig. 3c–f). This will result in a somewhat short measurement of hunt distance in those cases (apparent qualitatively in Supplementary Video 2).

GPS–INS processing was used to reduce noise and improve precision in the position and velocity solution (Supplementary Fig. 3), as well as increasing the temporal resolution of the data. It also allowed determination of orientation, which is otherwise not directly measured. Because the GPS receiver also records raw pseudorange, Doppler and carrier phase measurements for each satellite, future data processing may use a stationary reference station to calculate a more accurate differential GPS solution. Use of a tightly coupled GPS–INS solution may also provide increased accuracy and robustness, especially during periods when a reduced number of satellites are tracked (for example, turns).

Extraction of parameters for analysis: speed, distance and stride timing. Stride timings for data cutting and stride frequency were determined from the axis of accelerometer aligned approximately in the cranio-caudal direction. These accelerations were first low-pass filtered at twice anticipated stride frequency (8 Hz), and a peak detection algorithm was used to detect forward acceleration peaks at minimum duration of 0.2-s apart (equal to a maximum stride frequency of 5 Hz).

Horizontal speed was calculated from filtered velocity and averaged over the calculated strides (v_i) to remove the effects of speed fluctuation through the stride and collar oscillation relative to the centre of mass. These data were then smoothed with a rolling average (see below). Run distance was calculated by zero-order hold integration of the stride averaged horizontal speeds over the duration of the run. Maximum speed during each run was determined from these values. Stride frequency was calculated from the duration between stride timing peaks. For consistency in comparison, other parameters were then determined using the same method as in ref. 3, using only two-dimensional position and speed measurements. Position data were first down-sampled to the calculated stride times. The displacement vectors between consecutive positions were then calculated:

$$\overrightarrow{P_{i-1}P_i} = \overrightarrow{P_i} - \overrightarrow{P_{i-1}}$$

and

$$\overrightarrow{P_iP_{i+1}} = \overrightarrow{P_{i+1}} - \overrightarrow{P_i}$$

in which $\overrightarrow{P_i}$ is the two-dimensional position at sample/stride i .

Extraction of parameters for analysis: acceleration and power. A signed change of heading ($\Delta\theta_i$), and hence heading angular velocity (ω_i), were then calculated from the angle between the two vectors:

$$\Delta\theta_i = \sin^{-1} \left(\frac{\left| \overrightarrow{P_iP_{i+1}} \times \overrightarrow{P_{i-1}P_i} \right|}{\left| \overrightarrow{P_iP_{i+1}} \right| \left| \overrightarrow{P_{i-1}P_i} \right|} \right)$$

and

$$\omega_i = \frac{\Delta\theta_i}{\Delta T}$$

in which ΔT is the sampling interval.

The tangential or forward acceleration ($a_{t,i}$) and centripetal acceleration ($a_{c,i}$), as well as instantaneous turn radius (r_i) were then calculated:

$$a_{t,i} = \frac{v_{i+1} - v_i}{\Delta T}$$

$$a_{c,i} = \frac{v_i^2}{r_i} = \omega_i v_i$$

$$r_i = \frac{v_i}{\omega_i}$$

Finally mass-specific COM power was calculated as the dot product of stride averaged acceleration and stride averaged velocity (that is, multiply forward acceleration by forward speed):

$$k_i = a_i v_i$$

Mass-specific COM stride work (net COM kinetic energy change in a stride) was calculated as change in speed over a stride multiplied by stride average speed.

Extraction of parameters for analysis: improving accuracy through averaging. One important consideration when calculating heading, change of heading, and heading angular velocity from position measurements is that accuracy will decrease as speed decreases. Although averaging over a stride and across strides markedly improves the accuracy, lower average speed values will still be less accurate. The noise present is of a level that does not unduly influence extreme values even at very low speeds.

Although validations carried out on the stride timing show that it is generally accurate (Supplementary Fig. 2f), detection of an incorrect or spurious peak for end of stride would result in one stride duration being under or overestimated, and the adjacent stride duration being affected in the opposite manner. This would introduce error in parameters that do not change smoothly through a stride, such as acceleration and kinetic energy. We therefore applied a weighted average in which the stride period was averaged, with the mean of the duration of the preceeding and following stride. The weighted average was of the form:

$$S_{i,w} = 0.5S_{i-1} + S_i + 0.5S_{i+1}$$

in which S represents the parameter being weighted, and i is the stride number.

This approach was used as follows: tangential acceleration and hence acceleration power were calculated based on a weighted average stride speed. Centripetal acceleration was based on weighted stride speed and weighted heading rate. Stride duration was also weighted. Where these parameters have been plotted against horizontal speed, the weighted stride speed was also used. Applying more averaging than this did not change the distribution of outliers to a discernible extent

(Supplementary Fig. 4), but applying no averaging did result in more outliers giving us confidence in our extreme values with this treatment.

Extraction of parameters for analysis: grip and manoeuvring. Maximum traction has been proposed as a potential constraint to turning performance³. Coefficient of friction, μ , is the maximum achievable ratio of horizontal force (acceleration) with respect to vertical force (acceleration). Average vertical force is equal to acceleration due to gravity and assuming that vertical and horizontal forces are always in proportion:

$$\mu \geq \frac{ma}{mg}$$

So that maximum horizontal force and horizontal acceleration (a) are:

$$ma_{\max} = \mu mg$$

$$a_{\max} = \mu g$$

in which g is acceleration due to gravity, and m is mass. Substituting for horizontal acceleration in terms of tangential (a_t) and centripetal components (a_c):

$$\sqrt{a_t^2 + a_c^2} = \mu g$$

This demonstrates the potential trade-off between tangential and centripetal accelerations. Given that maximum centripetal acceleration will occur at constant speed ($a_t = 0$), and likewise that maximum tangential acceleration will occur in a straight line ($a_c = 0$):

$$a_{c, \max} = \mu g$$

$$a_{t, \max} = \mu g$$

Remembering that centripetal acceleration:

$$a_c = \frac{v^2}{r}$$

in which v is horizontal speed, and r is radius of turn. We form an equation for maximum speed (v_{\max}) in terms of turn radius (r):

$$\frac{v_{\max}^2}{r} = \mu g$$

$$v_{\max} = \sqrt{\mu g r}$$

A maximum limit for tangential acceleration based on maximum available muscle power (K) is derived as follows. When force and velocity are in the same direction:

$$K = Fv$$

$$K = ma_t v$$

Where F is force magnitude, v is horizontal speed, a_t is tangential acceleration and m is body mass. Given specific power by body mass (k):

$$k = \frac{K}{m}$$

Substituting gives:

$$a_{t, \max} = \frac{k_{\max}}{v}$$

Geometric limit to acceleration. A pitch limit for acceleration was previously proposed²⁰ that assumes that propulsion is derived purely from hip extension. This gives an acceleration limit for greyhounds of 10 m s^{-2} at all speeds derived from back length and leg length, and the limit for cheetahs would be similar as body height and length are similar⁷. Such a limit is not exceeded in our data (Fig. 4b), but there are few low speed acceleration strides.

Collar validation. A lurcher (greyhound/whippet/terrier cross in this case) dog was fitted with a mark 2 collar and encouraged to undertake maximal accelerations and sharp running turns on a beach in England, UK (the dog was accustomed to collar-testing experiments). The position of each footfall was determined using Survey grade GPS (OEM4, Novatel). Dual frequency Doppler and pseudorange and phase GPS data were post-processed relative to a local base station data using Waypoint GrafNav 8.10 (Novatel) with a horizontal accuracy of 20 mm. The

timing of each footfall was determined from simultaneous high-speed video at 500 frames per second (f.p.s.) (X-Pri 1280 × 1024 AOS GmbH). The camera trigger event was captured via an interrupt channel on an RVC GPS logger module with sub-millisecond accuracy, and used to express footfall events in GPS time for comparison to collar data (Supplementary Fig. 2e). The four footfalls per stride were easily identified in the position data (Supplementary Fig. 2a, b), and the distance between subsequent non-lead forefootfalls was defined as stride length, and the time between those foot falls as stride duration. Stride duration by video and by processing of collar data was compared by subtracting stride time from foot falls on high-speed video from stride duration from collar data and plotting the difference as a histogram (Supplementary Fig. 2f). Speed was calculated by dividing stride length by stride duration, and data were smoothed with a three-stride centre weighted rolling average as described for the collar data and the results plotted (Supplementary Fig. 2d). These data show that qualitatively the collar reproduces the track of the footfalls and that the speed time (and hence acceleration) data are indistinguishable between the two approaches. Further trials and analysis are required for a full assessment of the two methods.

Statistics. To establish which aspects of a run correlate with success, GLMMs were performed in R statistical software (R, version 2.14.1, 2011. R Development Core Team 2011, Foundation for Statistical Computing, Vienna, Austria). In the model, all the descriptive parameters of each hunt (terrain, distance, top speed, peak acceleration and deceleration number of turns and total turn angle) were included as fixed effects. To control for individual variation, a subject was included as a random effect. If an effect was not significant, and removing it from the model improved the Akaike information criterion (AIC), then it was removed. A chi-squared test was used to evaluate the effect of terrain on outcome.

Human acceleration power. Ten-metre split times for the 9.58 s world 100-m record run by Usain Bolt in 2009 were retrieved from the IAAF website (<http://berlin.iaaf.org/records/biomechanics/index.html>). A fifth order polynomial was fitted through the distance–time data. This polynomial visually fitted the data points and was differentiated to give formulae for speed and acceleration through the race and a function for instantaneous power through the race calculated as the product of the functions for speed and acceleration. This gave a peak centre of mass power of 25 W kg^{-1} body mass at 7 m s^{-1} , which is similar to previously published values for human sprinters¹⁹.

Hunting, terrain and outcome (success). Runs were identified in activity summaries by very high-peak acceleration amplitudes in all three axes, but particularly high accelerations in the cranio-caudal direction were the best indicator, confirmed from GPS speed where present. If two run events were within 10 min of one another, they were considered to be the same event for outcome measures. Terrain was determined from Google Earth; georeferencing of known landmarks and road junctions was confirmed to be accurate to within 5 m in the study area.

We identified feeding as a consistent signal on all three accelerometer axes (mean amplitude similar to mean of mean amplitudes), with particularly low cranio-caudal accelerations (compared with walking) and no change in location. See ref. 13 for more discussion. We classified a run as a successful hunt if 6 min of this feeding behaviour occurred in the 30 min after a run was identified. These methods correctly identified nine out of the ten known successful hunts using only the activity data (that is, without using GPS data), and correctly identified all nine as successful hunts. When applied to the main data set, the classification outcome correlated to other markers of success in 97% of known hunts. The other markers were: prey struggling captured in the accelerometer signal; cheetah remaining at hunt location for over two hours after the run; observing the cheetah on a kill.

List of symbols. i , stride number; \vec{P}_i , two-dimensional position; $\Delta\theta_i$, signed change of heading; ω_i , heading angular velocity; ΔT , sampling interval; a_i , horizontal acceleration; $a_{t,i}$, tangential or forward acceleration; $a_{c,i}$, centripetal acceleration; r_i , instantaneous turn radius; v_i , stride averaged horizontal speed; K , whole-body power; k_i , mass-specific whole-body power; S_i , parameter to be weighted; $S_{i,w}$, parameter after weighting; μ , coefficient of friction; m , body mass; g , acceleration due to gravity.

30. Kalman, R. E. A new approach to linear filtering and prediction problem. *Trans. ASME J. Basic Engineering* **82D**, 34–45 (1960).
31. Titterton, D. H. & Weston, J. L. *Strapdown Inertial Navigation Technology* 2nd edn, Vol. 207, Chs 3 and 11 (AIAA, 2004).
32. Rauch, H. E., Tung, F. & Striebel, C. T. Maximum likelihood estimates of linear dynamic systems. *AIAA J.* **3**, 1445–1450 (1965).

X-ray structure of the mammalian GIRK2- $\beta\gamma$ G-protein complex

Matthew R. Whorton^{1,2} & Roderick MacKinnon^{1,2}

G-protein-gated inward rectifier K⁺ (GIRK) channels allow neurotransmitters, through G-protein-coupled receptor stimulation, to control cellular electrical excitability. In cardiac and neuronal cells this control regulates heart rate and neural circuit activity, respectively. Here we present the 3.5 Å resolution crystal structure of the mammalian GIRK2 channel in complex with $\beta\gamma$ G-protein subunits, the central signalling complex that links G-protein-coupled receptor stimulation to K⁺ channel activity. Short-range atomic and long-range electrostatic interactions stabilize four $\beta\gamma$ G-protein subunits at the interfaces between four K⁺ channel subunits, inducing a pre-open state of the channel. The pre-open state exhibits a conformation that is intermediate between the closed conformation and the open conformation of the constitutively active mutant. The resultant structural picture is compatible with ‘membrane delimited’ activation of GIRK channels by G proteins and the characteristic burst kinetics of channel gating. The structures also permit a conceptual understanding of how the signalling lipid phosphatidylinositol-4,5-bisphosphate (PIP₂) and intracellular Na⁺ ions participate in multi-ligand regulation of GIRK channels.

In 1921, Otto Loewi established the existence of chemical synaptic transmission by showing that vagus nerve stimulation slows the heart rate through release of a chemical substance he called *vagusstoff*^{1,2}. *Vagusstoff* was later shown to be acetylcholine, the major neurotransmitter of the parasympathetic nervous system^{2,3}. Once released from the vagus nerve, acetylcholine binds to the M2 muscarinic receptor, a G-protein-coupled receptor (GPCR) in heart cell membranes, and causes the release of G-protein subunits G α and G $\beta\gamma$ from the receptor's intracellular surface⁴. The G $\beta\gamma$ subunits activate GIRK channels, causing them to open^{5–10}. Open GIRK channels drive the membrane voltage towards the resting (Nernst K⁺) potential, which slows the rate of membrane depolarization, as depicted (Fig. 1a). In atrial pacemaker cells of the heart, this directly decreases firing frequency and thus heart rate¹¹. Isoforms of the GIRK channel also exist in neurons, which permit G-protein-mediated regulation of neuronal electrical excitability¹².

For several decades, electrophysiological and biochemical methods have been applied to understand how G-protein subunits activate GIRK channels. Specific mutations on the G $\beta\gamma$ subunit^{13–17} and on the channel^{18,19} were shown to alter G-protein-mediated activation of GIRK channels. Biochemical and NMR studies identified components of both the G protein and channel that appear to interact with each other^{20–22}. Together these studies point to a direct interaction between the G-protein subunits and the channel to achieve channel activation. Here we present the crystal structure of a GIRK channel bound to G $\beta\gamma$ subunits, a key signalling complex in the G-protein-mediated control of electrical excitability.

GIRK2 activation by G-protein subunits

Our study addresses mouse GIRK2 (Kir3.2; inward rectifier K⁺ channel (Kir)), a neuronal GIRK channel that is able to function as a tetramer of identical subunits²³. Activation of GIRK2, which hereafter we refer to as GIRK, by GPCR stimulation is shown using an assay in which the M2 muscarinic GPCR is co-expressed together with GIRK channels in *Xenopus laevis* oocytes²⁴ (Fig. 1b, left). Initial replacement of Na⁺ by K⁺ in the extracellular solution causes some current to flow into the

oocyte, measured using two-electrode voltage clamp. When acetylcholine is then applied, a larger inward K⁺ current is turned on. Inhibition of current by tertiapin-Q, a bee venom toxin derivative that is known to inhibit the GIRK channel but not endogenous oocyte channels, establishes the current as mediated by the GIRK channels, a fraction of which are active in the absence of acetylcholine²⁵. The fraction of current activated by acetylcholine is variable, depending on the oocyte. Isolated membrane patches show the characteristic gating of single GIRK channels (Fig. 1b, right). These channels display ‘burst kinetics’, during which time an activated channel flickers rapidly between conducting (open) and non-conducting (closed) states, a property we will consider later. These electrophysiological recordings and other functional studies were carried out with the same construct used for crystallization and structural analysis. Hereafter, we refer to this construct, which consists of residues 52–380, as the wild-type channel. We emphasize that removal of the disordered amino and carboxy termini does not appear to alter the functional properties of the channel in any of the electrophysiological and flux measurements we made.

All studies of G-protein-mediated GIRK channel activation to date have been carried out with native cells or with cell lines in which components were expressed heterologously (as in Fig. 1b). Having obtained individual isolated components—the GIRK channel, G $\beta\gamma$ subunits and the signalling lipid PIP₂—we tested whether these alone (that is, in the absence of other cellular components) are sufficient to produce a competent signalling complex. Using a flux assay in which the isolated components are reconstituted into synthetic lipid vesicles, we find that they are indeed sufficient: baseline K⁺ flux observed in the absence of G $\beta\gamma$ is strongly enhanced in the presence of G $\beta\gamma$ (Fig. 1c). GIRK is also activated by intracellular Na⁺, which accounts for the greater flux observed in Na⁺ than in N-methyl-D-glucamine (NMDG⁺) (refs 26–29). However, even in the presence of Na⁺, G $\beta\gamma$ still causes significant enhancement of flux. These measurements with purified, reconstituted components confirm the conclusion reached through electrophysiological studies, that G $\beta\gamma$ in the presence of membranes containing PIP₂ is sufficient to increase the open probability of GIRK

¹Laboratory of Molecular Neurobiology and Biophysics, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. ²Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.

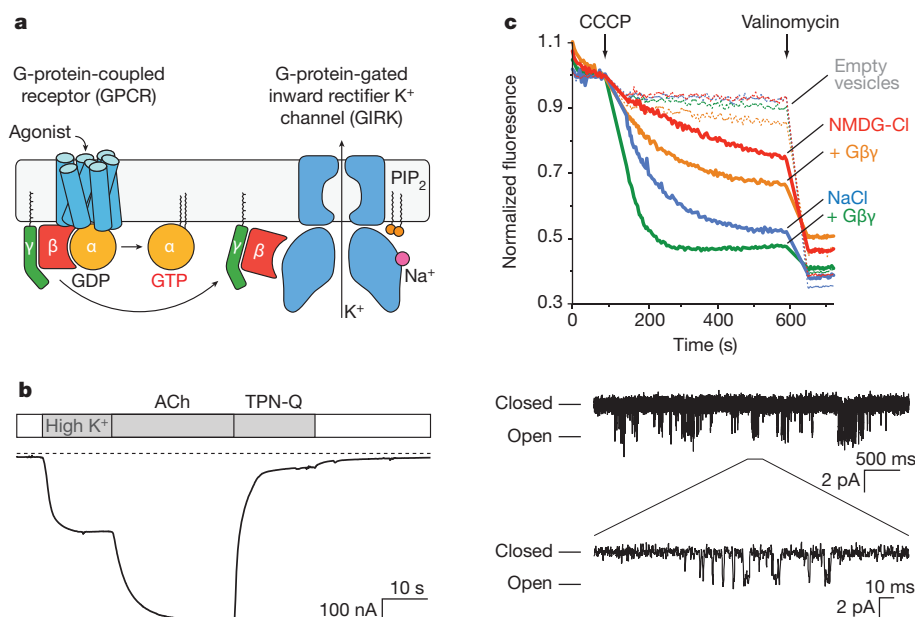


Figure 1 | Functional properties of the channel. **a**, Schematic of GPCR activation of GIRK channels. Agonist binding to a GPCR promotes the exchange of GDP for GTP on a bound G protein. This causes the G protein to dissociate from the receptor. The G α and G $\beta\gamma$ subunits subsequently dissociate from each other and they can then interact with effector proteins. G $\beta\gamma$ binding to the cytoplasmic domain of a GIRK channel in the presence of PIP₂ causes the channel to open. GIRK channels are also activated by elevated levels of intracellular Na⁺ ions. **b**, Example of GPCR-activation of GIRK. The truncated GIRK construct used for crystallography was co-expressed with the M2 muscarinic receptor in *X. laevis* oocytes. Whole-cell current was measured using two-electrode voltage clamp while holding the cell at -60 mV. The white bars indicate a physiological extracellular solution, whereas the grey bars represent a solution containing 98 mM KCl. The application of 10 μ M acetylcholine (ACh, a M2R agonist), or 1 μ M of tertiapin-Q (TPN-Q, a specific

channels. These experiments do not exclude a possible role for the G α subunit in regulating the GIRK channel or in conferring G-protein specificity, that is, explaining why GIRK normally is activated by G $\beta\gamma$ subunits associated with 'inhibitory' G $\alpha_{i/o}$ subunits and not by G $\beta\gamma$ subunits associated with stimulatory G α_s subunits⁸. Although many questions remain, these reconstitution experiments show that G $\beta\gamma$ by itself is sufficient to activate the GIRK channel.

Role of membrane in complex formation

Efforts to purify a stable GIRK-G $\beta\gamma$ protein complex in detergent solutions were unsuccessful. We therefore attempted to grow crystals of the complex in dodecylmaltoside by combining individually purified GIRK and G $\beta\gamma$ proteins at a twofold to threefold molar excess of G $\beta\gamma$ in the presence of a tenfold molar excess of PIP₂. Crystals containing both GIRK and G $\beta\gamma$ grew and diffracted to 3.5 Å resolution. These were of space group I422 with one GIRK monomer and one G $\beta\gamma$ complex per asymmetric unit. Phases were solved by molecular replacement using previously determined structures of GIRK and G $\beta\gamma$ as search models^{30,31}. A model of the complex was built and refined to working and free residuals, R_w and R_f , of 22.8% and 26.5%, respectively (Supplementary Table 1). The biological unit consists of one channel tetramer, four G $\beta\gamma$ subunits, four PIP₂ molecules and four Na⁺ ions bound to regulatory sites in addition to K⁺ ions in the selectivity filter. Intracellular Na⁺, PIP₂ and G $\beta\gamma$ are all physiological regulators of GIRK channel gating^{5–8,26–29,32}.

The arrangement of protein molecules within the crystal lattice is notable in light of a functional phenomenon known as 'membrane delimited' activation of GIRK channels by G-protein stimulation (Supplementary Fig. 1a, b). Electrophysiological studies showed that on their way to reaching the channel, G $\beta\gamma$ subunits behave as if to

GIRK2 blocker) is also indicated. The traces under the dashed line represent negative, inward currents. Single-channel recordings in the on-cell patch-clamp configuration (right). The patch pipette contained 96 mM KCl and 10 μ M ACh. The patch was held at -100 mV. A closeup of one of the burst openings is shown beneath. **c**, Activation of purified GIRK channels reconstituted into lipid vesicles. Channel activity was monitored in the presence of either NMDG-Cl or NaCl using a fluorescence-based assay described in detail in the Methods section. Purified G $\beta\gamma$ was added to some of the samples in either the NMDG-Cl or NaCl buffers, as indicated. The initiation of K⁺ flux by the addition of the H⁺ ionophore carbonyl cyanide m-chlorophenyl hydrazone (CCCP) is indicated. The addition of the K⁺ ionophore valinomycin to measure total flux capacity of the vesicles is also indicated. The dashed lines represent the same experimental conditions, except that the vesicles do not contain any GIRK.

diffuse while attached to the membrane's cytoplasmic surface (that is, membrane delimited)^{5–10}. In the crystal we observe pseudo-membrane layers consisting of transmembrane channel domains (TMDs) and aqueous layers consisting of cytoplasmic channel domains (CTCDs) and G $\beta\gamma$ subunits (Supplementary Fig. 1a, b). The G $\beta\gamma$ subunits are oriented such that the C terminus of the G γ subunit, which contains a covalent lipid molecule, a geranylgeranyl group, is pointed directly at the membrane layer as if to function as an anchor (Fig. 2a)^{33,34}. We note that a similar arrangement of the G-protein subunits was observed in the crystal of the β_2 -adrenergic GPCR in complex with the G $\alpha\beta\gamma$ heterotrimer, which was determined in lipid cubic phases³⁵. Thus, the GIRK-G $\beta\gamma$ crystal is compatible with physiological membrane delimited G $\beta\gamma$ activation of GIRK. We also note that our ability to achieve a complex in a crystal with membrane-like layers, but not in detergent solutions, implies that a membrane is important in the formation of the complex between GIRK and G $\beta\gamma$.

The protein complex

Two views of the GIRK channel show the gating regulators Na⁺, PIP₂ and G $\beta\gamma$ subunits bound (Fig. 2a, b). On the extracellular side of the membrane, the channel's turrets, which surround the pore entryway, project approximately 10 Å beyond the membrane surface. Previously we speculated that GIRK channels are more susceptible to pore-blocking toxins than are some other Kir channels, because the turrets are more widely spaced in GIRK and thus allow toxins to access the pore³¹. The structure of the turrets in this crystal are better defined than in the previous GIRK structures³¹, and indeed support this hypothesis (Supplementary Fig. 2).

On the intracellular side, large CTCDs project beyond the membrane surface, approximately 50 Å into the cell (Fig. 2a). These domains

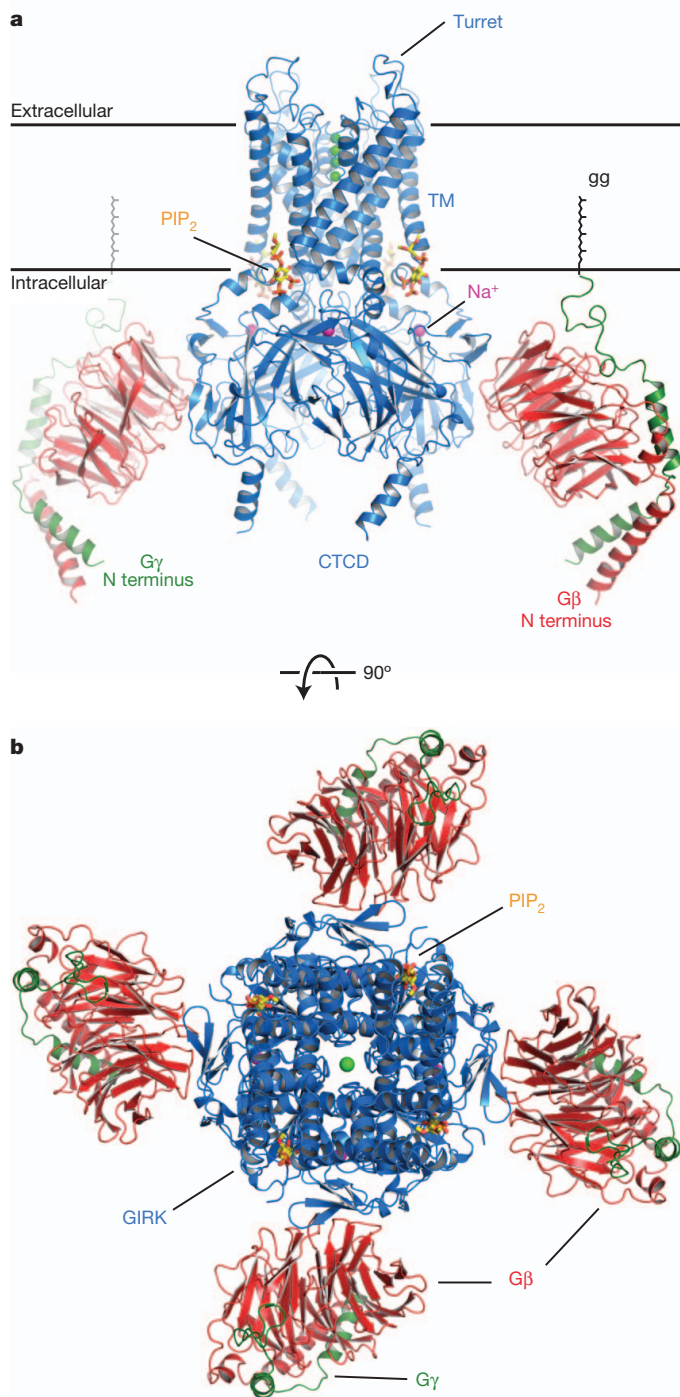


Figure 2 | Overall structure of the GIRK–G $\beta\gamma$ complex. **a**, A side view of the GIRK (blue), G β (red) and G γ (green) complex. The front G $\beta\gamma$ dimer was removed for clarity. The approximate extent of the phospholipid bilayer is shown by the black lines. The ‘gg’ label indicates the geranylgeranyl lipid modification at the C terminus of G γ . Bound Na⁺ ions are shown as purple spheres, the PIP₂ molecules are shown as sticks and the K⁺ ions as green spheres. **b**, Top-down view of the complex from the extracellular side of the cell.

provide an extensive surface through which molecules inside the cell can bind to regulate channel gating. The G $\beta\gamma$ subunits interact directly with the CTCDs through G β and to the membrane through the covalent lipid attached to G γ . Although we do not see the lipids in the crystal, the inferred covalent lipid interaction with the membrane is depicted. The entire GIRK channel–G $\beta\gamma$ complex forms a 120 Å × 120 Å square against the intracellular surface of the membrane (Fig. 2b).

The contact surface between GIRK and G β is approximately 700 Å² (Fig. 3a–c). On GIRK the contact surface is formed by secondary structure elements β K, β L, β M and β N from one channel subunit and by elements β D and β E from an adjacent channel subunit (Fig. 3d). The occurrence of the binding site at the interface between two channel subunits is likely to be important for mechanistic reasons, discussed below. On G β the contact surface is formed by β -sheet elements that form blades 1 and 7 on one edge of the β -propeller (Fig. 3c, e).

We compared the contact surfaces observed in the crystal structure with inferences drawn using other biophysical methods. Transferred cross saturation and chemical shift perturbation NMR experiments have been used to identify amino acids on the GIRK CTCDs that interact with G $\beta\gamma$ or change upon its binding²⁰. These amino acids, coloured purple (and orange for L344), fall mainly within or near the perimeter of the surface on GIRK that contacts G $\beta\gamma$ in the crystal structure (Supplementary Fig. 3a). Mutational studies also identified numerous amino acids that affect GIRK activation by G-protein stimulation^{13–19}. These amino acids are coloured orange on the surface of GIRK and G $\beta\gamma$ (Supplementary Fig. 3a, b). There are some outliers that may influence function indirectly or alternatively may disrupt protein structure, but most lie within or near the GIRK–G $\beta\gamma$ contact surface. Thus, both NMR and mutagenesis studies lend support to a biologically relevant signalling complex formed in the crystal structure.

Molecular determinants of G $\beta\gamma$ binding

All members of the Kir channel family share the same general molecular architecture, but, as far as we know, only the GIRKs (Kir3 proteins) are directly regulated by G-protein subunits³⁶. Many amino acids that compose the G $\beta\gamma$ binding surface on GIRK are also conserved among the G-protein-independent Kir channels, but a small set are unique to GIRK (Supplementary Fig. 4). This unique set includes Gln 248 and Phe 254 on the β D– β E loop and the sequence Leu-Thr/Ser-Leu (342–344) on the β L– β M loop (Fig. 3f). Gln 248 forms contacts with Gln 75, Ser 98 and Trp 99 on the G β subunit and mutations at Ser 98 and Trp 99 diminish G $\beta\gamma$ activation of GIRK (Fig. 3f–h)^{13,15}. The Leu-Thr/Ser-Leu sequence contacts Leu 55 and Lys 78 on G β (Fig. 3f, g, i). Mutations involving these residues also affect G $\beta\gamma$ activation of GIRK¹³. Thus, we can begin to understand G $\beta\gamma$ recognition in terms of short-range interactions afforded by a relatively small set of residues on the surface of the GIRK channel to which the G $\beta\gamma$ subunits bind.

Long-range electrostatic interactions between GIRK and G $\beta\gamma$ also appear significant (Fig. 4a, b). Several acidic residues on the β L– β M loop of the GIRK CTCDs complement an electropositive swath on the binding surface of the G β subunit (Fig. 4a, b, d). By comparison, a G-protein-independent Kir channel contains several lysine amino acids on its β L– β M loop that render its surface potential less electro-negative (Fig. 4c, e). Thus, electrostatic complementarity probably plays a role in binding affinity and specificity. In addition, by acting over longer distances, electrostatic forces are able to guide diffusing molecules into the formation of an encounter complex, where short-range interactions are then able to take hold³⁷. Such long-range guidance would seem to make sense here by directing the diffusion of G $\beta\gamma$ to the K⁺ channel once it is released from an activated GPCR.

Charged lipids on the membrane’s inner leaflet dominate the electric field close to the membrane where G-protein signalling occurs³⁸. With this in mind, Fig. 4f illustrates another potentially important role for electrostatic interactions between GIRK and G $\beta\gamma$. In the β_2 -adrenergic GPCR–G $\alpha\beta\gamma$ complex, the G $\beta\gamma$ subunits appear oriented to maximize positive protein charge contact with negative charges on the membrane surface³⁵. In the GIRK–G $\beta\gamma$ complex, although G $\beta\gamma$ resides at the same level with respect to the plane of the membrane, it is tilted roughly 35°. The tilt should reduce favourable electrostatic interactions between G $\beta\gamma$ and the membrane, but should produce new favourable interactions with GIRK to compensate. Thus, favourable electrostatic interactions between GIRK and G $\beta\gamma$ may help to reorient G $\beta\gamma$ with respect to the membrane.

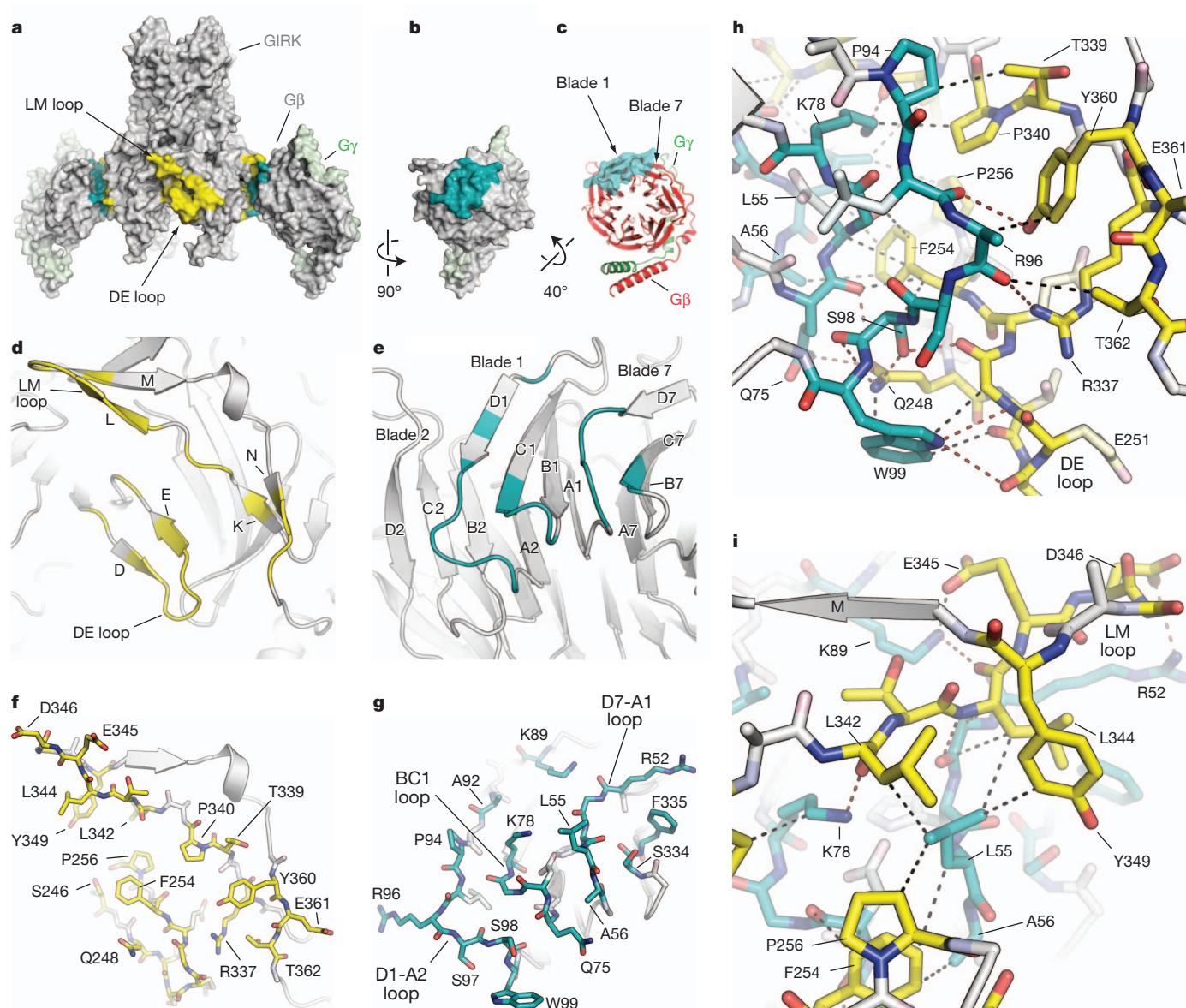


Figure 3 | The GIRK–Gβγ binding interface. **a**, Surface representation of the GIRK–Gβγ complex. The binding site on GIRK is coloured yellow and the binding site on Gβγ is coloured cyan. The front Gβγ dimer is removed for clarity. The overall orientation is the same as in Fig. 2a. **b**, A 90°-rotated view of a Gβγ dimer from panel **a** to more clearly show the binding interface. **c**, The Gβγ dimer is rotated upwards to orient the central axis of the β-propeller orthogonal to the page. **d**, **e**, A cartoon rendering of the binding interface on GIRK (**d**) and Gβ (**e**). Residues involved in the binding site are coloured yellow

on GIRK (**d**) and cyan on Gβ (**e**), and respectively correspond to the highlighted regions in panels **a** and **b**. In the **d**, **e** and **f**, **g** pairs of panels, the binding site can be approximately recapitulated by rotating each panel 90° towards each other, like making a sandwich. **f**, **g**, The same view as in panels **d** and **e**, with the residues involved in the binding site instead shown as sticks. **h**, **i**, A closeup of the GIRK–Gβγ interaction, focused on the DE loop, βK and βN region (**h**) or the LM loop region (**i**) of GIRK. Selected hydrogen bond and van der Waals interactions are shown as dashed lines as a visual aid.

Supplementary Fig. 5 compares the different contact surfaces that Gβγ uses to interact with other proteins, including Gα and four other effector and regulatory proteins, including GIRK. These comparisons support three conclusions. First, the β propeller of Gβ creates a large sticky surface that enables a multitude of unique interactions. Second, the GIRK binding site on Gβγ overlaps the Gα binding site. This observation, although anticipated, underscores the necessity of receptor activation and G-protein subunit dissociation (into Gα and Gβγ) to achieve channel activation^{4–8}. Third, the contact surface of RGS9 (Regulator of G protein signalling 9) on Gβ₅ is essentially non-overlapping with that of GIRK on Gβ_{1γ2}, although a conformational change would be required in RGS9 to allow it to bind to a GIRK–Gβγ complex³⁹. RGS9 in the nervous system suppresses the activity of opioid- and

dopamine-mediated G-protein signalling^{40–42}. Further studies will be needed to determine whether these signalling pathways intersect.

Gating control by Gβγ

What does the complex structure tell us about the regulation of GIRK channel gating by the Gβγ subunits? With the exception of the C-terminal half of Gγ, which is displaced by a crystal contact, Gβγ is structurally unchanged whether bound to Gα or to GIRK. The GIRK structure on the other hand is altered by the presence of Gβγ. Most notably the CTCD is rotated about the channel axis 4° anticlockwise (viewed from the membrane) relative to the TMD (Fig. 5a, b and Supplementary Video 1). The CTCD rotation is associated with an unwrapping and splaying of the right-handed bundle of four inner helices that

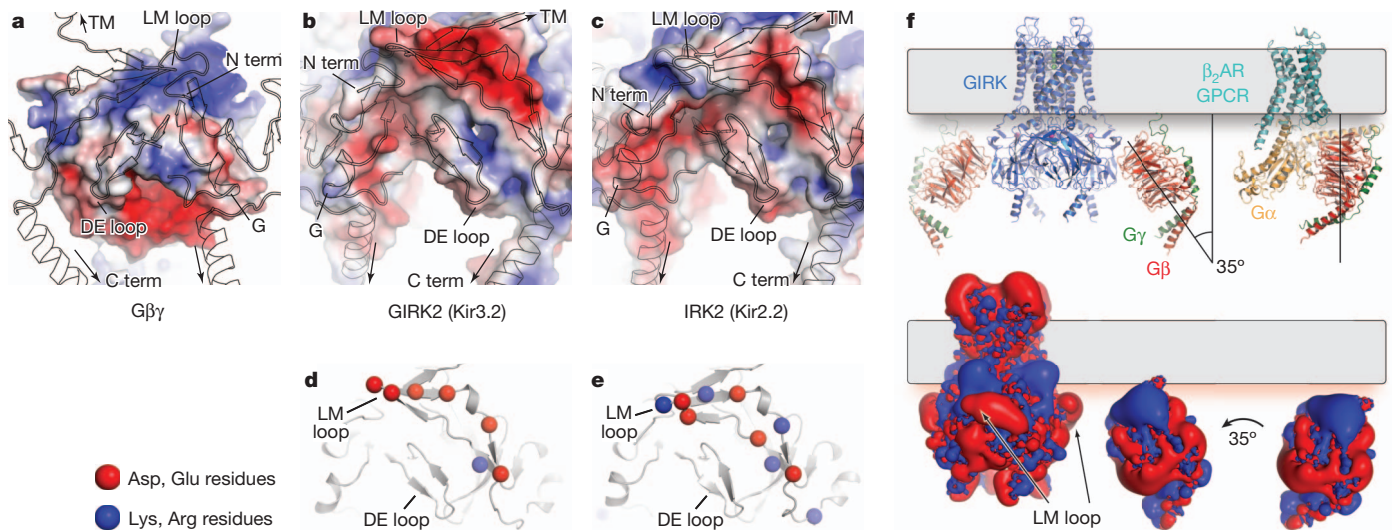


Figure 4 | The role of electrostatics at the GIRK-Gβγ interface. **a–c**, Surface representations are shown for Gβγ (**a**), GIRK2 (**b**) and IRK2 (ref. 47) (**c**), and are coloured according to calculated electrostatic surface potentials (−100 mV, red; +100 mV, blue). The proteins are shown in the same orientation as in Fig. 3d–g, except slightly zoomed out. A black outline of a cartoon representation of GIRK (or IRK2 in panel **c**) is overlaid to help the viewer match interacting surfaces. **d, e**, A closeup view of the LM loop region on GIRK2 (**d**) or

IRK2 (**e**). The Cα atom for aspartate or glutamate residues are shown as red balls, and arginines and lysines are shown as blue balls. **f**, Cartoon representations of the GIRK-Gβγ complex and the β₂-AR-Gαβγ complex³⁵ are shown. The black lines highlight the difference in the relative orientation of the two Gβγ dimers to the membrane (grey rectangle) (top). Isocontour representations of the electrostatic potential for an isolated GIRK channel (left) or a Gβγ dimer (middle and right) are shown (−25 mV, red; +25 mV, blue) (bottom).

form the closed ‘inner helix gate’ in the absence of Gβγ. Four Phe 192 side chains on the inner helices come together to form the narrowest constriction in the closed inner helix gate. These phenylalanine side chains are partially disordered in the slightly splayed structure with

Gβγ subunits bound, however, the pathway seems too narrow to conduct hydrated K⁺ ions: other clearly open K⁺ channels, such as the Kv1.2 voltage-dependent K⁺ channel, have a minimum diameter of 10 Å (distance between van der Waals surfaces), but in this structure of

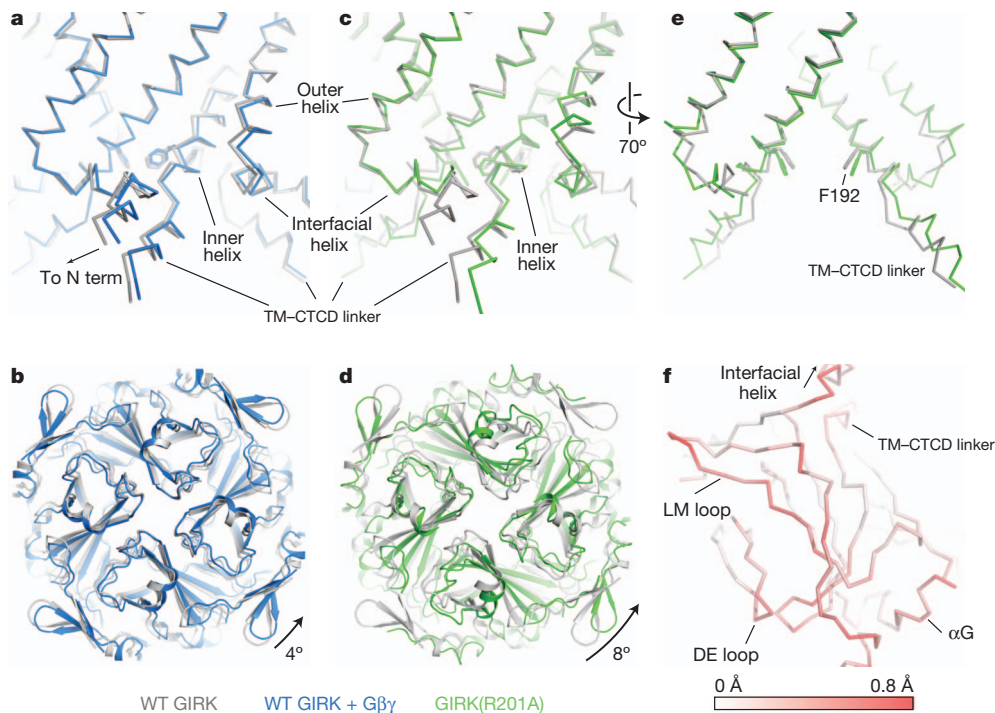


Figure 5 | Gβγ-induced conformational changes in GIRK. **a–e**, The structure of wild-type (WT) GIRK in complex with PIP₂ (Protein Data Bank (PDB): 3SYA) is shown (grey) and is compared to the wild-type GIRK-Gβγ complex (blue) and the GIRK(R210A) mutant in complex with PIP₂ (PDB: 3SYQ) (green). All structures are aligned by a structurally inert region around the selectivity filter at the top of the transmembrane domain to show the relative twisting of the cytoplasmic domains. Panels **a** and **c** show a Cα ribbon trace of a side view of the transmembrane domain, with Phe 192 shown as sticks for reference. Panels **b** and **d** show a top-down view of the cytoplasmic domain, with

the degree and direction of twisting indicated. Panel **e** is a 70° rotated view of panel **c** to highlight the conformational changes of the inner helices of the two subunits that bound PIP₂ in this structure. **f**, A Cα ribbon trace of part of the cytoplasmic domain from the wild-type GIRK plus Gβγ structure. Red colouring reflects the root mean squared deviation (r.m.s.d.) between the wild-type GIRK and wild-type GIRK plus Gβγ structures when they are aligned by their cytoplasmic domains. This highlights the additional conformational changes that happen in the cytoplasmic domain apart from the rigid-body twisting shown in panels **b** and **d**. The most intense red represents a r.m.s.d. of 0.8 Å.

GIRK2 it is only 6–7 Å^{43–46}. However, the single-channel recordings suggest perhaps we should not expect an open conformation in the crystal structure (Fig. 1b). G-protein subunits are most probably bound to the GIRK channel during the duration of an activity burst; however, during the burst the channel flickers rapidly with a relatively low open probability. This might suggest that the GIRK structure we observe in the presence of Gβγ, which adopts a distinctly different conformation than the structure without Gβγ, represents a G-protein-activated, pre-open conformation, corresponding to the channel part way along the reaction pathway from closed to open.

In a previous study, we determined the crystal structure of a constitutively-open point mutant of the channel, GIRK(R201A), which is conductive in the absence of G-protein stimulation³¹. The mutant structure is indeed open, and its comparison to the wild-type channels in the absence and presence of Gβγ is suggestive of a mechanism (Figs 5c–e and 6). In the mutant channel, the CTCD is rotated an additional 4° beyond the rotation caused by Gβγ, and the CTCD subunits have undergone an internal conformational change associated with widening of the membrane-facing apex of the CTCD. This widening further opens the inner helix gate to a diameter of 9 Å. One caveat is that only two PIP₂ molecules are bound to the tetramer in the mutant channel—to diagonally opposed subunits—so that opening is twofold rather than fourfold symmetric. Packing in the mutant crystal appears to have prevented the binding of PIP₂ molecules to all four subunits, which is observed in the wild-type structures³¹. We suspect that had four PIP₂ molecules bound to the mutant channel, then opening would be symmetric. Despite the asymmetry of GIRK(R201A), the conformation of GIRK in the Gβγ complex is clearly intermediate between the closed (Gβγ-free) and opened GIRK(R201A) structures. A morph between these conformations shows that binding of Gβγ causes a 4° rotation of the CTCDs and a slight splaying of the inner helices. The GIRK(R201A) mutation produces a further 4° rotation, a conformational change

within the CTCD subunits, and an opening of the inner helix gate (Supplementary Video 2). These conformations could account for the burst kinetic behaviour of single GIRK channels if binding of the Gβγ subunits produces a pre-open conformation in the membrane, from which the channel flickers rapidly between open (conductive) and pre-open (nonconductive) conformations (Fig. 6, highlighted pathway). This hypothesis would predict that the GIRK(R201A) mutant channels should exhibit a higher open probability. Unfortunately, due to reduced expression levels, we have been unable to characterize the single channel behaviour of this mutant in either *X. laevis* oocytes or Chinese hamster ovary cells.

Discussion

GIRK2 channels are regulated by PIP₂, G-protein subunits and intracellular Na⁺ ions^{5–8,26–29,32}. We show in reconstitution experiments, using purified components, that these regulators individually activate the channel partially, and in combination activate it to a greater extent. We present a crystal structure of GIRK with all three regulators bound. Together with previously determined crystal structures of GIRK and an R201A mutant of GIRK, both determined in the presence and absence of PIP₂, we have pieced together a structural description of conformational states that might underlie the sequential activation of GIRK channels (Fig. 6, highlighted pathway). The binding of Gβγ subunits to GIRK causes a rotation of the CTCDs with respect to the TMD and a partial splaying of the inner helices. This conformation is intermediate between the closed and GIRK(R201A) open conformations. Full opening is associated with a further rotation of the CTCDs and splaying open of the inner helical gate.

Together the structures permit conceptual explanations for multi-ligand regulation. PIP₂ is required for full gate opening in the GIRK (R201A) mutant channel³¹. Thus, PIP₂ seems to play a facilitative role; under conditions that favour opening, PIP₂ helps, presumably by

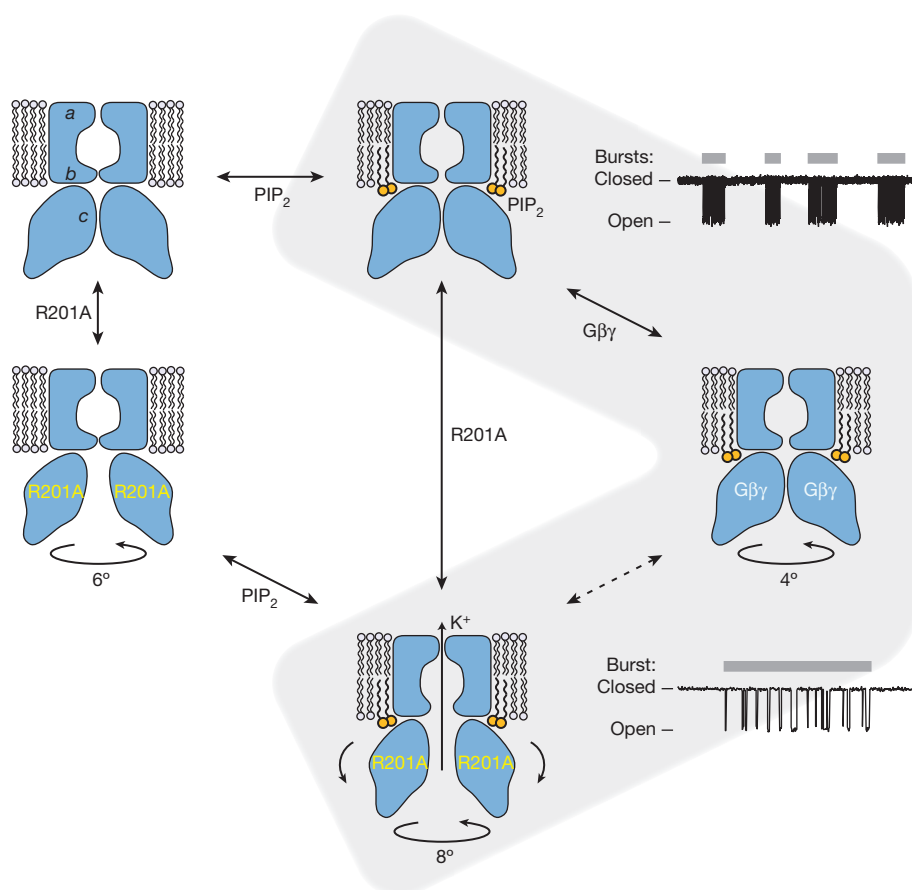


Figure 6 | A model of gating regulation of GIRK channels. The blue shapes depict a GIRK channel with a selectivity filter (a) and two gates: the inner helix gate (b) and the G loop gate (c). The structures correspond to wild-type GIRK without PIP₂ or Gβγ (PDB: 3SYO), wild-type GIRK with PIP₂ only (PDB: 3SYA), GIRK(R201A) without PIP₂ or Gβγ (PDB: 3SYP), wild-type GIRK with PIP₂ and Gβγ (PDB: 4KFM), and GIRK(R201A) with PIP₂ (PDB: 3SYQ). Circular arrows with degrees indicate CTCD rotation about the pore axis with respect to the TMD, relative to wild-type structures without Gβγ. Curved arrows in GIRK(R201A) with PIP₂ reflect the outward rocking of CTCD subunits observed in this structure. Idealized single-channel recordings are shown on the right (expanded time scale on the bottom) to illustrate our current hypothesis regarding the gating transitions that the channel undergoes. Inter-burst periods correspond to a channel with only PIP₂ bound (top); bursts (grey bars) correspond to the channel with PIP₂ and Gβγ bound, which fluctuates rapidly (indicated by dashed arrows) between non-conducting (right) and conducting (bottom) conformations.

strengthening the interface between the CTCD and TMD where PIP₂ is bound. The Na⁺ ion is bound to the CTCD at a position that undergoes a conformational change when the channel opens. Thus, we should expect Na⁺ binding to be thermodynamically coupled to channel gating, allowing Na⁺ to function as a regulator.

Concerning stoichiometry, we do not know how many Gβγ subunits are required to open the GIRK channel, but based on the structure we are compelled to speculate. Gβγ binding causes a rotation of the CTCD associated with splaying of the inner helices to open the gate. The rotation no doubt occurs because Gβγ binds at the interface between two adjacent CTCD subunits, which produces detectable relative motions of the subunits, and inferred strain between them (Fig. 5f and Supplementary Video 3). This, we believe, is the source of the rotation. One Gβγ subunit causing strain across one of four interfaces is probably not enough. Four is undoubtedly better. Can a single GPCR near a GIRK channel catalyse a sufficient number GDP to GTP exchange reactions and release Gβγ subunits to activate the channel or do several GPCRs surround a GIRK channel, and if the latter, are GIRK channels and GPCRs randomly distributed or organized in a stoichiometric cluster or array? The GIRK–Gβγ complex provides a starting point for addressing these questions.

METHODS SUMMARY

A truncated GIRK2 construct (containing residues 52–380) was expressed and purified from *Pichia pastoris* cells as previously described³¹, except that the detergent dodecyl-β-D-maltopyranoside (DDM) was used instead of decyl-β-D-maltopyranoside (DM). Full-length human Gβ₁ and Gγ₂ were co-expressed in High Five insect cells, extracted with cholate, purified by Talon metal affinity chromatography in anion exchange 3–12, and then further purified by ion exchange and size exclusion chromatography in CHAPS before adding 20% glycerol and freezing in liquid nitrogen. Aliquots were thawed as needed and the detergent was exchanged to DDM while bound to a HiTrap ion exchange column. Individually purified GIRK and Gβγ were combined along with C8-PIP₂ at 200 μM, 400 μM and 2 mM respectively and allowed to incubate for 2 h at room temperature before proceeding with crystal trials. Crystals of the complex were grown by the hanging drop vapour diffusion method at 20 °C using 600 mM NaK tartrate and 50 mM ADA pH 5.7–5.9 as the crystallization solution. Crystals were cryo-protected by briefly transferring to solutions containing 2.35–2.4 M NaK tartrate. The structure was solved by molecular replacement, using previously determined structures of both GIRK³¹ and Gβγ³⁰ as search models. The resulting model was improved by iterative rounds of refinement and manual adjustment. Electrophysiology recordings of GIRK channel function were performed by expressing GIRK in *X. laevis* oocytes and using two-electrode voltage clamp to measure whole cell currents, or on-cell patch clamp to measure single-channel activity. *In vitro* assessment of GIRK function was performed by reconstituting purified GIRK into lipid vesicles and then measuring GIRK activity under a variety of conditions using a fluorescence-based assay.

Full Methods and any associated references are available in the online version of the paper.

Received 12 February; accepted 3 May 2013.

Published online 5 June 2013.

- Loewi, O. Über humorale übertragbarkeit der herznervenwirkung. *Pflügers Arch.* **189**, 239–242 (1921).
- Loewi, O. The chemical transmission of nerve action (1936 Nobel lecture) in *Nobel Lectures in Physiology or Medicine 1922–1941* (World Scientific Publishing, 1999).
- Loewi, O. & Navratil, E. Über humorale übertragbarkeit der herznervenwirkung. *Pflügers Arch.* **214**, 678–688 (1926).
- Gilman, A. G. G proteins: transducers of receptor-generated signals. *Annu. Rev. Biochem.* **56**, 615–649 (1987).
- Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E. J. & Clapham, D. E. The βγ subunits of GTP-binding proteins activate the muscarinic K⁺ channel in heart. *Nature* **325**, 321–326 (1987).
- Reuveny, E. et al. Activation of the cloned muscarinic potassium channel by G protein βγ subunits. *Nature* **370**, 143–146 (1994).
- Wickman, K. D. et al. Recombinant G-protein βγ-subunits activate the muscarinic-gated atrial potassium channel. *Nature* **368**, 255–257 (1994).
- Pfaffinger, P. J., Martin, J. M., Hunter, D. D., Nathanson, N. M. & Hille, B. GTP-binding proteins couple cardiac muscarinic receptors to a K channel. *Nature* **317**, 536–538 (1985).

- Breitwieser, G. E. & Szabo, G. Uncoupling of cardiac muscarinic and β-adrenergic receptors from ion channels by a guanine nucleotide analogue. *Nature* **317**, 538–540 (1985).
- Soejima, M. & Noma, A. Mode of regulation of the ACh-sensitive K-channel by the muscarinic receptor in rabbit atrial cells. *Pflügers Arch.* **400**, 424–431 (1984).
- Trautwein, W. & Dudel, J. Zum mechanismus der membranwirkung des acetylcholin an der herzmuskelfaser. *Pflügers Arch.* **266**, 324–334 (1958).
- Lüscher, C. & Slesinger, P. A. Emerging roles for G protein-gated inwardly rectifying potassium (GIRK) channels in health and disease. *Nature Rev. Neurosci.* **11**, 301–315 (2010).
- Ford, C. E. et al. Molecular basis for interactions of G protein βγ subunits with effectors. *Science* **280**, 1271–1274 (1998).
- Albsoul-Younes, A. M. et al. Interaction sites of the G protein β subunit with brain G protein-coupled inward rectifier K⁺ channel. *J. Biol. Chem.* **276**, 12712–12717 (2001).
- Mirshahi, T., Robillard, L., Zhang, H., Hebert, T. E. & Logothetis, D. E. Gβ residues that do not interact with Gα underlie agonist-independent activity of K⁺ channels. *J. Biol. Chem.* **277**, 7348–7355 (2002).
- Zhao, Q. et al. Interaction of G protein β subunit with inward rectifier K⁺ channel Kir3. *Mol. Pharmacol.* **64**, 1085–1091 (2003).
- Zhao, Q. et al. Dominant negative effects of a Gβ mutant on G-protein coupled inward rectifier K⁺ channel. *FEBS Lett.* **580**, 3879–3882 (2006).
- He, C., Zhang, H., Mirshahi, T. & Logothetis, D. E. Identification of a potassium channel site that interacts with G protein βγ subunits to mediate agonist-induced signaling. *J. Biol. Chem.* **274**, 12517–12524 (1999).
- Finley, M., Arrabit, C., Fowler, C., Suen, K. F. & Slesinger, P. A. βL-βM loop in the C-terminal domain of G protein-activated inwardly rectifying K⁺ channels is important for Gβγ subunit activation. *J. Physiol. (Lond.)* **555**, 643–657 (2004).
- Yokogawa, M., Osawa, M., Takeuchi, K., Mase, Y. & Shimada, I. NMR analyses of the Gβγ binding and conformational rearrangements of the cytoplasmic pore of G protein-activated inwardly rectifying potassium channel 1 (GIRK1). *J. Biol. Chem.* **286**, 2215–2223 (2011).
- Slesinger, P. A., Reuveny, E., Jan, Y. N. & Jan, L. Y. Identification of structural elements involved in G protein gating of the GIRK1 potassium channel. *Neuron* **15**, 1145–1156 (1995).
- Huang, C. L., Slesinger, P. A., Casey, P. J., Jan, Y. N. & Jan, L. Y. Evidence that direct binding of Gβγ to the GIRK1 G protein-gated inwardly rectifying K⁺ channel is important for channel activation. *Neuron* **15**, 1133–1143 (1995).
- Kofuji, P., Davidson, N. & Lester, H. A. Evidence that neuronal G-protein-gated inwardly rectifying K⁺ channels are activated by Gβγ subunits and function as heteromultimers. *Proc. Natl Acad. Sci. USA* **92**, 6542–6546 (1995).
- Kubo, Y., Reuveny, E., Slesinger, P. A., Jan, Y. N. & Jan, L. Y. Primary structure and functional expression of a rat G-protein-coupled muscarinic potassium channel. *Nature* **364**, 802–806 (1993).
- Jin, W. & Lu, Z. Synthesis of a stable form of teritapin: a high-affinity inhibitor for inward-rectifier K⁺ channels. *Biochemistry* **38**, 14286–14293 (1999).
- Ho, I. H. & Murrell-Lagnado, R. D. Molecular mechanism for sodium-dependent activation of G protein-gated K⁺ channels. *J. Physiol. (Lond.)* **520**, 645–651 (1999).
- Ho, I. H. & Murrell-Lagnado, R. D. Molecular determinants for sodium-dependent activation of G protein-gated K⁺ channels. *J. Biol. Chem.* **274**, 8639–8648 (1999).
- Sui, J. L., Petit-Jacques, J. & Logothetis, D. E. Activation of the atrial K_{ACh} channel by the βγ subunits of G proteins or intracellular Na⁺ ions depends on the presence of phosphatidylinositol phosphates. *Proc. Natl Acad. Sci. USA* **95**, 1307–1312 (1998).
- Sui, J. L., Chan, K. W. & Logothetis, D. E. Na⁺ activation of the muscarinic K⁺ channel by a G-protein-independent mechanism. *J. Gen. Physiol.* **108**, 381–391 (1996).
- Wall, M. A. et al. The structure of the G protein heterotrimer G₁₂β₁γ₂. *Cell* **83**, 1047–1058 (1995).
- Whorton, M. R. & MacKinnon, R. Crystal structure of the mammalian GIRK2 K⁺ channel and gating regulation by G proteins, PIP₂, and sodium. *Cell* **147**, 199–208 (2011).
- Huang, C. L., Feng, S. & Hilgemann, D. W. Direct activation of inward rectifier potassium channels by PIP₂ and its stabilization by Gβγ. *Nature* **391**, 803–806 (1998).
- Mumby, S. M., Casey, P. J., Gilman, A. G., Gutowski, S. & Sternweis, P. C. G protein gamma subunits contain a 20-carbon isoprenoid. *Proc. Natl Acad. Sci. USA* **87**, 5873–5877 (1990).
- Yamane, H. K. et al. Brain G protein γ subunits contain an all-trans-geranylgeranylcytostine methyl ester at their carboxyl termini. *Proc. Natl Acad. Sci. USA* **87**, 5868–5872 (1990).
- Rasmussen, S. G. et al. Crystal structure of the β₂ adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
- Hibino, H. et al. Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiol. Rev.* **90**, 291–366 (2010).
- Sheinerman, F. B., Norel, R. & Honig, B. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153–159 (2000).
- McLaughlin, S. The electrostatic properties of membranes. *Annu. Rev. Biophys. Chem.* **18**, 113–136 (1989).
- Cheever, M. L. et al. Crystal structure of the multifunctional Gβ5–RGS9 complex. *Nature Struct. Biol.* **15**, 155–162 (2008).
- Zachariou, V. et al. Essential role for RGS9 in opiate action. *Proc. Natl Acad. Sci. USA* **100**, 13656–13661 (2003).
- Rahman, Z. et al. RGS9 modulates dopamine signaling in the basal ganglia. *Neuron* **38**, 941–952 (2003).
- Kovoor, A. et al. D2 dopamine receptors colocalize regulator of G-protein signaling 9–2 (RGS9–2) via the RGS9 DEP domain, and RGS9 knock-out mice develop

- dyskinesias associated with dopamine pathways. *J. Neurosci.* **25**, 2157–2165 (2005).
43. Jiang, Y. *et al.* X-ray structure of a voltage-dependent K⁺ channel. *Nature* **423**, 33–41 (2003).
 44. Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
 45. Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
 46. Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* **309**, 897–903 (2005).
 47. Hansen, S. B., Tao, X. & MacKinnon, R. Structural basis of PIP₂ activation of the classical inward rectifier K⁺ channel Kir2.2. *Nature* **477**, 495–498 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Hoff and members of D. Gadsby's laboratory (Rockefeller University) for assistance with oocyte preparation; Y. Hsiung for assistance

with insect cell culture; R. Sanishvili, N. Venugopalan, and S. Corcoran (GM/CA, Advanced Photon Source, Argonne National laboratory) for assistance at the synchrotron; and members of the MacKinnon laboratory. The use of the Rigaku/MSC microMax 007HF and Formulatrix robot in the Rockefeller University Structural Biology Resource Center was made possible by Grant Numbers 1S10RR022321-01 and 1S10RR027037-01, respectively, from the National Center for Research Resources of the National Institutes of Health (NIH). R.M. is an investigator in the Howard Hughes Medical Institute.

Author Contributions M.R.W. performed the experiments. M.R.W. and R.M. analysed the data and wrote the paper.

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited into the Protein Data Bank under accession code 4KFM. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M. (mackinn@rockefeller.edu).

METHODS

Molecular Biology. A truncated GIRK2 construct (containing residues 52–380) was cloned into the pPICZ, or pGEM vectors for high-level expression or electrophysiology, respectively, as previously described³¹. The full-length human G-protein β_1 and γ_2 subunits were cloned into pFastbac vectors. The γ_2 construct also included an N-terminal His10 tag, followed by a yellow fluorescent protein (YFP), and then a PreScission protease site (LEVLFQ/GP). Individual baculoviruses were made from these pFastbac vectors using the Bac-to-Bac system (Invitrogen).

Protein expression and purification. GIRK2 was expressed in *P. pastoris* as previously described³¹. GIRK2 was extracted and purified from *P. pastoris* cells essentially as previously described³¹, with a few exceptions: Dodecyl- β -D-maltopyranoside (DDM) was used in all steps instead of decyl- β -D-maltopyranoside (DM). DDM was used at 4% for extraction, 0.4% during the Talon purification, and 0.05% on the Superdex-200 column. 10 mM imidazole was included during the batch binding to the Talon resin. PreScission protease-cleaved protein was loaded onto the Superdex-200 column at a sufficiently high concentration such that the 1 ml peak was at least 1 mg ml⁻¹. This was done to reduce the final detergent concentration in the concentrated protein, which was necessary for growing large, thick crystals. The protein was concentrated to 30–40 mg ml⁻¹ in a 50K MWCO, and was typically used immediately.

The G proteins were expressed in High Five insect cells (Invitrogen). High Five cells were grown at 27 °C in Express Five serum-free media (Invitrogen), supplemented with L-glutamine. The cells were grown to a density of 1–2 million cells per ml and then infected with a volume of baculovirus for each protein empirically determined to give optimal expression (~30 ml). After 48 h, the cells were harvested by centrifugation at 4,000g for 15 min. The cell pellets were resuspended in a small volume of the supernatant and this slurry was transferred to 50 ml conical tubes (approximately one litre of cells per 50 ml tube). After another centrifugation at 4,000g for 15 min, the supernatant was removed and the cell pellets (~15 ml) were frozen in liquid N₂ and then stored at -80 °C until needed.

A typical G $\beta\gamma$ preparation involved purifying protein from 8 litres worth of cells. All procedures were performed at 4 °C unless indicated. Frozen cell pellets were added to 480 ml of room temperature buffer comprised of 50 mM HEPES, pH 8; 65 mM NaCl; 1 mM EDTA; 5 mM β -mercaptoethanol (β ME) and protease inhibitors (0.1 mg ml⁻¹ pepstatin, 1 mg ml⁻¹ leupeptin, 1 mg ml⁻¹ aprotinin, 0.1 mg ml⁻¹ soy trypsin inhibitor, 1 mM benzamide and 1 mM phenylmethylsulfonyl fluoride). This solution was stirred at room temperature in a stainless steel beaker until the pellets melted. Then the beaker was transferred to ice and the solution was sonicated for six times 1 min each using a probe sonicator (Branson), with a 1 min cool down in-between. The lysed cells were then spun at 35,000 g for 35 min to pellet the membranes. The supernatant was poured off and the pellets were each briefly rinsed with ~5 ml of 50 mM HEPES, pH 8; 50 mM NaCl; 0.1 mM MgCl₂; 5 mM β ME and protease inhibitors. The pellets were then resuspended in the same buffer using a dounce homogenizer to a final volume of 350 ml. Na-cholate was added to a final concentration of 1.5%, and the solution was stirred for 40 min. The solubilized membranes were spun again at 35,000 g for 35 min to pellet insoluble material. The supernatant was diluted with two volumes of dilution buffer (20 mM HEPES, pH 8; 300 mM NaCl; 5 mM β ME; 7.5 mM imidazole; 0.5% anzerger 3-12 (Anatrace) and protease inhibitors) and then added to 32 ml of Talon resin (Clontech) pre-equilibrated in dilution buffer. This suspension was stirred for 1 h, then spun at 1,000 g for 5 min. The resin was transferred to a column and washed with 4 column volumes (cv.) of 20 mM HEPES, pH 8; 300 mM NaCl; 5 mM β ME; 5 mM imidazole and 0.5% anzerger 3-12; then, 4 cv. of 20 mM HEPES, pH 8; 50 mM NaCl; 5 mM β ME; 10 mM imidazole and 0.5% anzerger 3-12. Then the protein was eluted from the column with 20 mM HEPES, pH 8; 40 mM NaCl; 5 mM β ME; 200 mM imidazole and 0.5% anzerger 3-12. Dithiothreitol (DTT) and EDTA were added to 5 mM and 1 mM, respectively. PreScission protease was also added at 1:20 protease:total protein, and incubated overnight. The next day, an additional amount of PreScission (1:40 protease:total protein) was added and incubated at room temperature for 2 h. This solution was then diluted down to a conductivity of ~5 mS cm⁻¹ with 20 mM HEPES, pH 8; 5 mM β ME; 1% anzerger 3-12. A white precipitate usually formed at this step, which was mostly comprised of contaminants. This was pelleted by a brief centrifugation, and the supernatant was further filtered through a 0.22- μ m filter before loading onto a Mono Q 16/10 column, equilibrated with buffer A (20 mM HEPES, pH 8; 40 mM NaCl; 5 mM β ME and 0.7% 3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulphonate/N,N-dimethyl-3-sulpho-N-[3-[[3 α ,5 β ,7 α ,12 α]-3,7,12-trihydroxy-24-oxocholan-24-yl]amino]propyl]-1-propanaminium hydroxide (CHAPS; Anatrace). The column was washed with 15 cv. buffer A, then the protein was eluted with a 50 cv. gradient from 0–20% buffer B (buffer A with 1 M NaCl). The G $\beta\gamma$ protein eluted as a major peak as well as several minor peaks that were assumed to be unprenylated or differentially phosphorylated species.

The major peak was collected and concentrated in a 30K MWCO concentrator to at least 5 mg ml⁻¹. This was then run on a Superdex-200 column equilibrated in 20 mM Tris, pH 7.5; 100 mM KCl; 5 mM DTT; 0.7% CHAPS) in multiple runs of ~2.5 mg protein per run. This helped to remove trace smaller molecular weight G $\beta\gamma$ protein, which was assumed to be an unprenylated species. Peak fractions were again concentrated in a 30K MWCO concentrator to ~10 mg ml⁻¹. Glycerol was added to 20% and the protein was frozen in liquid N₂ in 150 μ l aliquots and stored at -80 °C until needed. Approximately 8–10 mg worth of these aliquots were thawed as needed and the detergent was then exchanged to DDM while bound to a 1 ml Q Sepharose column (HiTrap, GE Healthcare) at room temperature. This was done to ensure complete detergent exchange as well as to get a high protein concentration without a high detergent concentration, which was necessary for growing large, thick crystals. DDM (anagrade) was added to 1% final from a 10% stock. This was then slowly diluted with two volumes of 20 mM Tris, pH 7.5; 5 mM DTT and 0.05% DDM, and then loaded onto the HiTrap column equilibrated with buffer C (20 mM Tris, pH 7.5; 30 mM KCl; 5 mM DTT and 0.05% DDM). The column was washed with 25 ml buffer C, then 3 ml 10% buffer D (buffer C with 300 mM KCl), 3 ml 20% buffer D, 9 ml 30% buffer D, then the protein was eluted with 100% buffer D. The middle of this elution peak yielded about 0.75 ml of 8–10 mg ml⁻¹. This was then further concentrated in a 30K MWCO concentrator. After one centrifugation spin of the concentrator, the protein was diluted with the appropriate volume of 20 mM Tris, 5 mM DTT, including an appropriate concentration of EDTA to bring the KCl concentration down to 150 mM and the EDTA concentration up to 1 mM. The protein was then further concentrated to 40–50 mg ml⁻¹ and stored on ice until needed.

Crystallization. A typical crystallization experiment involved mixing concentrated GIRK2, G $\beta\gamma$, and 1,2-dioctanoyl-sn-glycero-3-phospho-(1'-myo-inositol-4',5'-biphosphate) (C8-PIP₂, Avanti Polar Lipids) at a final concentration of 200 μ M, 400 μ M, and 2 mM, respectively. The mixture was incubated at room temperature for at least 2 h before mixing 1:1 (0.2 μ l + 0.2 μ l) with the crystallization solution (600 mM NaK tartrate, 50 mM Na-ADA (N-(2-acetamido)iminodiacetic acid) pH 5.7–5.9). The crystals were grown at 20 °C using the hanging drop vapour diffusion method. The crystals appeared after a few days and grew as thin square plates or plate clusters to full size within a week. The crystals were cryoprotected by first adding 1 μ l of a solution containing 20 mM Tris, pH 7.5; 150 mM KCl; 1 mM EDTA; 0.5% DDM; 1 mM DTT; 720 mM NaK tartrate; 50 mM Na-ADA, pH 5.7 and 1 mM C8-PIP₂ directly to the drop. Crystals were gently broken off of the clusters and then briefly transferred to a new solution containing 20 mM Tris, pH 7.5; 0.5% DDM; 1 mM DTT; 50 mM Na-ADA, pH 5.7; 1 mM C8-PIP₂ and 2.35–2.4 M NaK tartrate, depending on the crystal size. The crystals were then flash-frozen in liquid N₂.

Structure determination. Diffraction data were collected at the Advanced Photon Source 23ID-B beamline ($\lambda = 1.033$ Å). Diffraction images were processed with the HKL2000 program suite⁴⁸. The crystals diffracted anisotropically (3.45 \times 3.45 \times 3.8 Å along the a*, b*, and c* axes, respectively), so integrated diffraction data were truncated to these diffraction limits using a script available from the UCLA Diffraction Anisotropy Server⁴⁹. The crystals were highly sensitive to radiation damage, so data from three sites on one crystal and one site on another crystal were scaled together in HKL2000 to form the most complete data set. R_{merge} and R_{pim} diffraction data statistics were calculated using the RMERGE program⁵⁰.

The structure was solved with Phaser⁵¹ by sequentially searching for a GIRK2 monomer (PDB: 3SYA) and then a G $\beta_1\gamma_2$ dimer (PDB: 1GP2, chains B and G). Initial rigid-body refinement of the molecular replacement solution with REFMAC^{52,53} identified the twist between the cytoplasmic and transmembrane domains of GIRK2. The model was then further modified in Coot⁵⁴ and refined with REFMAC, using jelly-body restraints. The translation/libration/screw (TLS) refinement in REFMAC was used in the final rounds of refinement.

Because of the poor quality of electron density in the K⁺ selectivity filter, distance restraints were used during refinement between the K⁺ ions in the selectivity filter to constrain their positions based on known properties from high-resolution K⁺ channel structures. In the final stages of refinement, a strong electron density feature near the interfacial helix of the channel was modelled as a DDM maltose headgroup with a five-carbon aliphatic chain. This was based on bilobal shape of the density, the location of this density at the presumed boundary of the DDM micelle, and the presence of ≥ 15 mM DDM in the crystallization condition. PIP₂ and Na⁺ ligands were carried over from the original search model and their presence was confirmed using simulated annealing omit maps (Supplementary Fig. 6a, b).

Comprehensive model validation was performed with MolProbity⁵⁵ (as embedded within PHENIX⁵⁶), with 94.5/5.5% of residues falling within the favoured and allowed region of the Ramachandran plot, respectively. Simulated annealing omit maps were calculated using PHENIX. Data collection and refinement statistics are shown in Supplementary Table 1. All figures were made using PyMOL (<http://www.pymol.org>). Videos were made in PyMOL using intermediate structures

interpolated with the CNS^{57,58} script from the Yale Morph Server^{59,60}. Electrostatics were calculated using APBS and visualized using the APBS plugin in PyMOL⁶¹. Disordered side chains were added back to the model in the most common rotamer conformation to make the calculations more accurate.

Electrophysiology. Two-electrode voltage-clamp recordings of GIRK2 currents in *X. laevis* oocytes were performed as previously described³¹. For patch-clamp recordings of GIRK2 currents, *X. laevis* oocytes were injected with complementary RNA for the M2 muscarinic receptor and GIRK2 as previously described. The patch pipettes typically had a resistance of 4 MΩ and were filled with 96 mM KCl; 1.8 mM CaCl₂; 1 mM MgCl₂; 10 mM K-HEPES, pH 7.5 and 1 μM acetylcholine. The bath solution contained 96 mM KCl, 5 mM EGTA and 10 mM K-HEPES, pH 7.5. The recordings were made in the on-cell configuration and the patch was held at −100 mV. The currents were recorded with an Axon Axopatch 200B (Molecular Devices), filtered at 1 kHz and sampled at 10 kHz using an analogue-to-digital converter (Digidata 1440A, Axon Instruments). pClamp10.1 software (Axon Instruments) was used for controlling the amplifier and data acquisition. Uninjected oocytes showed no detectable currents.

Reconstitution into lipid vesicles. Purified GIRK2 was reconstituted into lipid vesicles by first mixing chloroform solutions of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG), and 1-α-phosphatidylinositol-4,5-bisphosphate (PI(4,5)P₂); from porcine brain; predominant acyl chains are 18:0 and 20:4) at mass ratios of 3:1:0.04, and then drying this solution under an argon gas stream. The dried lipid film was placed in a vacuum desiccator for a few hours and then resuspended in 20 mM K-HEPES, pH 7.35; 150 mM KCl; 1 mM EDTA; 35 mM CHAPS (at 10 mg ml^{−1}) by incubating for 2 h at room temperature with periodic sonication. Purified GIRK2 was concentrated to ~2 mg ml^{−1} and then added to 100 μl of the solubilized lipids, typically at a 1:300 protein:lipid mass ratio, and incubated at room temperature for 30 min. Dehydrated 1 ml Sephadex G-50 (Sigma) columns were prepared by loading 1 ml of hydrated Sephadex G-50 resin (equilibrated in 20 mM K-HEPES, pH 7.35; 150 mM KCl; 1 mM EDTA) on to a small plastic spin column, and then briefly centrifuging at 1,500g. The protein-lipid mixture was then gently pipetted onto the resin bed and the columns were briefly spun up to 1,000g to remove the detergent and form the proteoliposomes.

Flux assay. 10 μl of vesicles were added to 190 μl of flux buffer (20 mM K-HEPES, pH 7.35; 150 mM NaCl; 1 mM EDTA and 2 μM of the pH-sensitive dye 9-Amino-6-chloro-2-methoxyacridine (ACMA)) in a 96-well black plate (in some experiments, the NaCl was replaced with N-methyl-D-glucamine HCl (NMDG-Cl)). This creates a K⁺ gradient across the vesicle membrane and a negative potential inside the vesicle if there are open K⁺ channels. Fluorescence was monitored

every 5 s at 410/490 nm (excitation/emission, 20 nm bandwidth). After a baseline was established, the H⁺ ionophore carbonyl cyanide m-chlorophenyl hydrazone (CCCP) was added to 1 μM from a 400 μM stock and briefly mixed. This allows protons to enter the vesicles, drawn in by the negative potential, which causes quenching of ACMA. After 10 min, the K⁺ ionophore valinomycin is added to 20 nM from a 8 μM stock and briefly mixed. This acts as a shunt to indicate the total capacity of the vesicles. All fluorescence time courses were normalized to the fluorescence before CCCP addition to account for slight differences in fluorescence between the wells. In some experiments, a concentrated stock of CHAPS-solubilized purified Gβγ was added to the vesicles before diluting them in the flux buffer to give a final concentration of 180 nM.

48. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (ed. Carter, C. W. Jr) 307–326 (Academic Press, 1997).
49. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
50. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34**, 130–135 (2001).
51. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
52. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
53. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
54. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
55. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
56. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
57. Brünger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
58. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
59. Echols, N., Milburn, D. & Gerstein, M. MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.* **31**, 478–482 (2003).
60. Krebs, W. G. & Gerstein, M. The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res.* **28**, 1665–1675 (2000).
61. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).

Sodium content as a predictor of the advanced evolution of globular cluster stars

Simon W. Campbell¹, Valentina D'Orazi^{1,2}, David Yong³, Thomas N. Constantino¹, John C. Lattanzio¹, Richard J. Stancliffe^{3,4}, George C. Angelou¹, Elizabeth C. Wylie-de Boer³ & Frank Grundahl⁵

The asymptotic giant branch (AGB) phase is the final stage of nuclear burning for low-mass stars. Although Milky Way globular clusters are now known to harbour (at least) two generations of stars^{1,2}, they still provide relatively homogeneous samples of stars that are used to constrain stellar evolution theory^{3–5}. It is predicted by stellar models that the majority of cluster stars with masses around the current turn-off mass (that is, the mass of the stars that are currently leaving the main sequence phase) will evolve through the AGB phase^{6,7}. Here we report that all of the second-generation stars in the globular cluster NGC 6752—70 per cent of the cluster population—fail to reach the AGB phase. Through spectroscopic abundance measurements, we found that every AGB star in our sample has a low sodium abundance, indicating that they are exclusively first-generation stars. This implies that many clusters cannot reliably be used for star counts to test stellar evolution timescales if the AGB population is included. We have no clear explanation for this observation.

We obtained high-resolution spectra ($R \approx 24,000$) for a sample of 20 AGB stars and 24 red giant branch stars in the Galactic globular cluster NGC 6752. The spectral coverage included the strong Na I doublet at 5,880 Å. In Fig. 1 we show the stellar sample. We include red giant branch stars as a control group, because it has previously been shown that this evolutionary population harbours the standard abundance distributions, including the well-known Na–O anticorrelation present in all globular clusters⁸.

Our sodium abundance results are shown in Fig. 2. The red giant branch sample shows the usual spread in $[Na/Fe] = \log_{10}(N_{Na}/N_{Fe})_{\text{star}} - \log_{10}(N_{Na}/N_{Fe})_{\text{Sun}}$ of roughly 1 dex (N_x is the number density of atoms of each elemental species). On the other hand, the AGB result is striking—all the AGB stars in our sample lie at the low end of the red giant branch distribution. The upper envelope of the AGB sodium abundances is located at about $[Na/Fe] = 0.18$ dex. This corresponds very closely to a previous red giant branch study that defines the Na-poor, first-generation population as having $[Na/Fe] \lesssim 0.2$ dex (their ‘Primordial’ population)¹. We find the proportion of Na-poor to Na-rich red giant branch stars in our data to be 30:70. This also corresponds well to the roughly 30:70 proportions found previously¹. Thus all of our AGB stars appear to be first-generation stars, giving a first generation to second generation ratio change from 30:70 in the red giant branch population to 100:0 in the AGB population. The range in $[Na/Fe]$ in our AGB sample is very small, with a mean of -0.07 dex and a standard deviation of 0.10 dex. This scatter is comparable to our internal uncertainties (Fig. 2 and Supplementary Table 1), which suggests that the AGB stars may have a uniform abundance of Na.

Our results indicate that the entire population of second-generation stars, having increased levels of Na, must fail to enter the AGB phase. This is a very significant effect, because the second-generation population contains the majority of the stars in NGC 6752 (70%). Although two studies have theorized that some AGB stars may not ascend the

AGB in NGC 6752 (refs 9, 10), this has been based on low-resolution cyanogen band strength observations. It is known that these observations are affected by many uncertainties¹¹, including the *in situ* variation of C and N along the red giant branch¹². Measurement of elemental Na is more robust, because Na is not affected by molecular band formation uncertainties and stars of these masses (~ 0.8 solar masses) cannot alter their Na abundances *in situ*. In particular, a reduction in surface Na would require very high temperatures at the base of the convective envelope. This is only achieved in much more massive stars ($\gtrsim 4$ solar masses), via ‘hot bottom burning’ nucleosynthesis¹³. Thus the result presented here is to our knowledge the first conclusive confirmation that stars of certain chemical composition do not ascend the AGB. Moreover, we can readily identify which stars avoid the AGB on the basis of their Na content.

An obvious consequence of such a large proportion of stars avoiding the AGB is that there should currently be many fewer stars in the AGB phase than expected. A detailed study reporting star counts of globular cluster populations finds a value of $R_2 = N_{\text{AGB}}/N_{\text{HB}} \approx 0.06$ for NGC 6752 (ref. 14; N_{AGB} and N_{HB} are the number of AGB and horizontal branch stars, respectively). This is one of the lowest R_2 values in their globular cluster sample. The globular clusters with the two highest R_2 values in their sample (M 5 and M 55) could be assumed to provide an upper limit to R_2 , because R_2 is fairly insensitive to metallicity, He abundance and globular cluster age¹⁵. This upper value is roughly 0.18, that is, a factor of 3 higher than that of NGC 6752. This is indeed

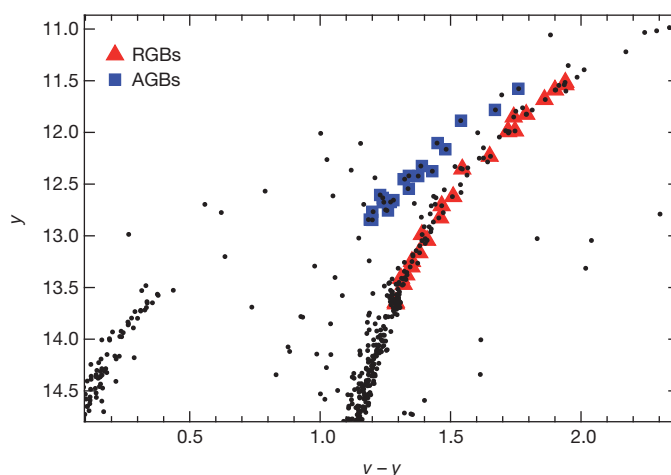


Figure 1 | Sample selection in the Strömgren *uvby* colour-magnitude diagram of NGC 6752. Small black dots show the whole photometric sample¹⁸. Our AGB and red giant branch (RGB) stellar samples are shown as blue squares and red triangles respectively. Part of the horizontal branch can be seen at bottom left, at y magnitudes $\gtrsim 13.5$.

¹Monash Centre for Astrophysics, School of Mathematical Sciences, Monash University, Melbourne, Victoria 3800, Australia. ²Department of Physics and Astronomy, Macquarie University, Balaclava Road, North Ryde, Sydney, New South Wales 2109, Australia. ³Research School of Astronomy and Astrophysics, Australian National University, Weston, Australian Capital Territory 2611, Australia. ⁴Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany. ⁵Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 1, DK-8000 Aarhus C, Denmark.

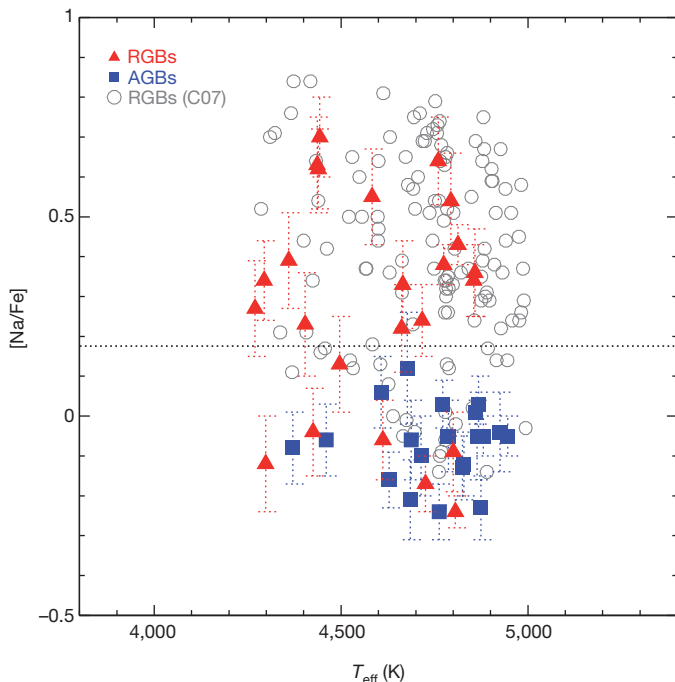


Figure 2 | Sodium abundance results for NGC 6752. $[\text{Na}/\text{Fe}]$ for our sample of red giant branch stars (red triangles, 24 stars) and AGB stars (blue squares, 20 stars) shown against stellar effective temperature T_{eff} (also see Supplementary Table 1). For comparison, the red giant branch results of a previous study (C07)⁸ are included (grey open circles). The horizontal dotted line at $[\text{Na}/\text{Fe}] = 0.18$ marks the upper envelope of AGB values, and divides Na-rich and Na-poor stars. The spectroscopic observations were carried out with the Very Large Telescope. The FLAMES/Giraffe HR11 grating was used, with a spectral coverage $\lambda = 5,597\text{--}5,840\text{ \AA}$. Na abundances (assuming local thermodynamic equilibrium) were obtained from the strong Na I doublet at $5,680\text{ \AA}$, with the driver *ab find* in MOOG²⁰ (2011 version, available at <http://www.as.utexas.edu/~chris/moog.html>) and the Kurucz model atmospheres with no overshooting²¹. Stellar parameters were derived in the following way: T_{eff} values were calculated from a Strömgren colour $(b-y)$ calibration²²; gravities were then computed from stellar luminosities and the derived temperatures (with assumed stellar mass M of 0.8 solar masses, and distance modulus of $(m-M)_V = 13.30$), while microturbulence values ξ were computed using a relation from the literature²³. A metallicity of $[\text{Fe}/\text{H}] = -1.54$ dex (ref. 8) was adopted for all stars. Although the lines under scrutiny are known to be not largely affected by departures from local thermodynamic equilibrium at these metallicities and temperatures (at most 0.15 dex), we applied non local thermodynamic equilibrium corrections²⁴ to our Na abundances. Error bars show the random (internal) uncertainties (see also Supplementary Table 1). The uncertainties were estimated by adding, in quadrature, errors due to the equivalent width measurements and those related to stellar parameters. The latter were evaluated in the standard way, by varying one parameter at a time and inspecting the corresponding variation in the resulting abundances. We adopted errors of $\Delta T_{\text{eff}} = \pm 30\text{ K}$, $\Delta \log g = 0.1$, $\Delta \xi = 0.1\text{ km s}^{-1}$, and $\Delta[\text{Fe}/\text{H}] = \pm 0.05\text{ dex}$.

consistent with our result of $\geq 70\%$ of stars not ascending the AGB. Current stellar model predictions for R_2 tend to be lower than 0.18, being around 0.12–0.15 (refs 15, 16); however, the models are known to suffer from significant uncertainties¹⁶. We note that the observed R_2 value for NGC 6752 is still at least a factor of 2 smaller than the model predictions.

Studies of the stage of evolution directly preceding the AGB, the horizontal branch phase, can shed light on the AGB avoidance phenomenon. One recent investigation into the Na abundances in a sample of horizontal branch stars in NGC 6752 showed that the redder end of the horizontal branch (NGC 6752 only has a blue horizontal branch) contains only Na-poor stars¹⁷ (see Supplementary Information section 2 for more discussion). This implies that it is the bluer (presumably Na-rich) horizontal branch stars that must avoid AGB ascent, leaving only the redder, Na-poor horizontal branch stars to populate the AGB

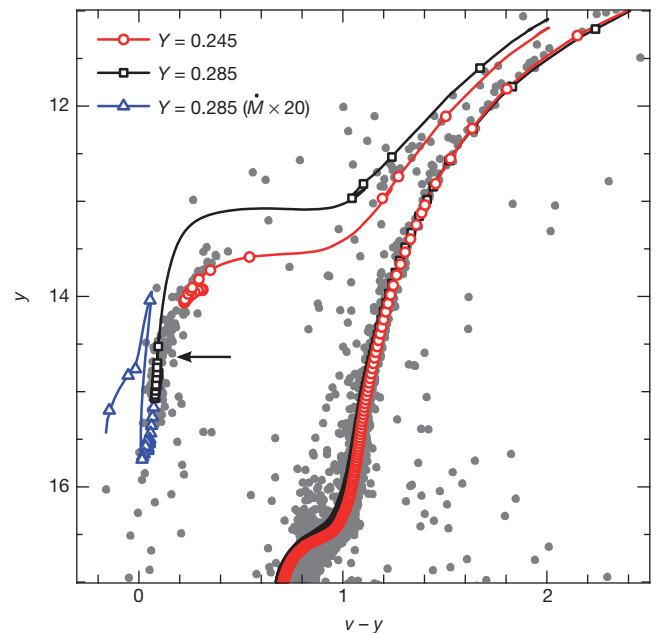


Figure 3 | Theoretical stellar model tracks overlain on the Strömgren colour-magnitude diagram of NGC 6752. The solid red line (with open circles marking 5 Myr time intervals) is a model with an initial mass of 0.8 solar masses and a helium content of $Y = 0.245$. This Y value matches that reported for the redder end of the horizontal branch¹⁷. This first-generation model does indeed spend most of its horizontal branch evolution at the redder end of the horizontal branch. The solid black line (with open squares marking 5 Myr time intervals) is a model with an initial mass of 0.75 solar masses and an enhanced helium content of $Y = 0.285$. This second-generation model spends its horizontal branch evolution in a bluer part of the horizontal branch, but still ascends the AGB, contrary to the observational findings of the current study. The solid blue line (with open triangles marking 5 Myr time intervals) shows the evolution of the $Y = 0.285$ model with an *ad hoc* 20-fold increase in mass loss rate ($\dot{M} = dM/dt$) initiated once the star settles on the horizontal branch. This model evolves downwards along the extreme blue end of the horizontal branch and fails to ascend the AGB. The arrow indicates the location of the Grundahl jump at $y = 14.65$ (see text for details). The stellar models were calculated using the Monash University stellar structure code MONSTAR²⁵. The code has been recently updated with low-temperature opacity tables which follow variations in C, N and O²⁶. The usual Reimers mass loss rate²⁷ was used for the red giant branch and horizontal branch (with $\eta = 0.48$ for the models with normal mass-loss). Transformations from the theoretical luminosity- T_{eff} plane to the colour-magnitude diagram have been made²⁸.

of NGC 6752. Combining this information with our ratio of Na-poor to Na-rich stars (Fig. 2), we can estimate the horizontal branch colour at which stars fail to ascend the AGB. We find an ‘ascension cut-off’ colour that coincides exactly with the colour for the ‘Grundahl jump’. The Grundahl jump is a discontinuity in horizontal branch morphology which is seen in all globular clusters studied to date whose horizontal branch extends beyond an effective temperature of $\sim 11,500\text{ K}$ (refs 18, 19). Thus it appears that all stars bluer than the Grundahl jump do not ascend the AGB, at least in NGC 6752. This may represent further evidence that there is some fundamental change in the stellar atmosphere structure and/or mass-loss physics occurring at the Grundahl jump temperature¹⁸.

We have calculated some stellar models to compare with the photometric observations. The model results are shown in Fig. 3. The model representative of the first-generation stars populates the redder end of the horizontal branch, before the Grundahl jump, as expected. It then continues to the AGB. The model representative of the He-rich second-generation stars (presumed to correspond to the Na-rich population) populates the bluer end of the horizontal branch (after the Grundahl jump), and also continues to the AGB. Thus our second-generation model cannot account for the lack of ascension of the

Na-rich blue horizontal branch stars in this part of the colour-magnitude diagram. We note that an increased mass loss rate during the red giant branch phase would result in a bluer zero-age horizontal branch star (see Supplementary Information section 3 for more detail), so this can not be a solution, because the colour-magnitude diagram is clearly populated in this region. We speculate that one solution to this problem may be that the horizontal branch stars blueward of the Grundahl jump experience enhanced mass loss. To test this, we artificially increased the mass loss rate in the second-generation model during its horizontal branch phase by an *ad hoc* factor of 20 (Fig. 3). Indeed this model can populate the blue end of the horizontal branch and also fail to become an AGB star. We note however that this test is purely hypothetical and, although this result may provide a starting point, a thorough investigation into the reasons behind the discordance between theory and observation is sorely needed. There is currently no clear explanation for such a high proportion of failed AGB stars.

Finally, because globular clusters are often used to test stellar evolution theory, the extremely high AGB failure rate reported here will affect any test which uses star counts of AGB stars. This is true of the *R* method used to check the lifetimes of various phases of evolution. In particular, the R' , R_1 and R_2 values^{3,4,14} all involve the number of AGB stars, so these values will be flawed (including the globular cluster He values inferred from them). This is particularly important if the globular clusters in question have blue extensions to their horizontal branches, because it is the blue horizontal branch stars that appear not to ascend the AGB. Star number counts used to ascertain AGB lifetimes will also be misleading, unless the proportion of AGB ascenders is known somehow—for example, via an ascension cut-off in Na abundance or horizontal branch colour.

Received 4 February; accepted 12 April 2013.

Published online 29 May 2013.

- Carretta, E., Bragaglia, A., Gratton, R. & Lucatello, S. Na-O anticorrelation and HB. VIII. Proton-capture elements and metallicities in 17 globular clusters from UVES spectra. *Astron. Astrophys.* **505**, 139–155 (2009).
- Gratton, R. G., Carretta, E. & Bragaglia, A. Multiple populations in globular clusters. Lessons learned from the Milky Way globular clusters. *Astron. Astrophys. Rev.* **20**, 50 (2012).
- Iben, I. & Rood, R. T. Ratio of horizontal branch stars to red giant stars in globular clusters. *Nature* **224**, 1006–1008 (1969).
- Buonanno, R., Corsi, C. E. & Fusi Pecci, F. The giant, asymptotic, and horizontal branches of globular clusters. II — Photographic photometry of the metal-poor clusters M15, M92, and NGC 5466. *Astron. Astrophys.* **145**, 97–117 (1985).
- Renzini, A. & Fusi Pecci, F. Tests of evolutionary sequences using color-magnitude diagrams of globular clusters. *Annu. Rev. Astron. Astrophys.* **26**, 199–244 (1988).
- Kippenhahn, R. & Weigert, A. *Stellar Structure and Evolution* (Springer, 1990).
- Landsman, W. B. *et al.* Ultraviolet imagery of NGC 6752: a test of extreme horizontal branch models. *Astrophys. J.* **472**, L93–L96 (1996).
- Carretta, E., Bragaglia, A., Gratton, R. G., Lucatello, S. & Momany, Y. Na-O anticorrelation and horizontal branches. II. The Na-O anticorrelation in the globular cluster NGC 6752. *Astron. Astrophys.* **464**, 927–937 (2007).
- Norris, J., Cottrell, P. L., Freeman, K. C. & Da Costa, G. S. The abundance spread in the giants of NGC 6752. *Astrophys. J.* **244**, 205–220 (1981).
- Campbell, S. W. *et al.* The case of the disappearing CN-strong AGB stars in Galactic globular clusters — preliminary results. *Mem. Soc. Astron. Ital.* **81**, 1004 (2010).
- Yong, D., Grundahl, F., Johnson, J. A. & Asplund, M. Nitrogen abundances in giant stars of the globular cluster NGC 6752. *Astrophys. J.* **684**, 1159–1169 (2008).
- Smith, G. H. & Tout, C. A. The production of surface carbon depletions among globular cluster giants by interior mixing. *Mon. Not. R. Astron. Soc.* **256**, 449–456 (1992).
- Boothroyd, A. I., Sackmann, I.-J. & Ahern, S. C. Prevention of high-luminosity carbon stars by hot bottom burning. *Astrophys. J.* **416**, 762–768 (1993).
- Sandquist, E. L. & Bolte, M. Exploring the upper red giant and asymptotic giant branches: the globular cluster M5. *Astrophys. J.* **611**, 323–337 (2004).
- Cassisi, S., Salaris, M. & Irwin, A. W. The initial helium content of galactic globular cluster stars from the R-parameter: comparison with the cosmic microwave background constraint. *Astrophys. J.* **588**, 862–870 (2003).
- Cassisi, S. & Castellani, V. Degl'Innocenti, S., Piotto, G. & Salaris, M. asymptotic giant branch predictions: theoretical uncertainties. *Astron. Astrophys.* **366**, 578–584 (2001).
- Villanova, S., Piotto, G. & Gratton, R. G. The helium content of globular clusters: light element abundance correlations and HB morphology. I. NGC 6752. *Astron. Astrophys.* **499**, 755–763 (2009).
- Grundahl, F., Catelan, M., Landsman, W. B., Stetson, P. B. & Andersen, M. I. Hot horizontal-branch stars: the ubiquitous nature of the “jump” in Strömgren u , low gravities, and the role of radiative levitation of metals. *Astrophys. J.* **524**, 242–261 (1999).
- Momany, Y. *et al.* A new feature along the extended blue horizontal branch of NGC 6752. *Astrophys. J.* **576**, L65–L68 (2002).
- Snedden, C. A. *Carbon and Nitrogen Abundances in Metal-Poor Stars*. Ph.D. thesis, Univ. Texas at Austin (1973).
- Kurucz, R. ATLAS9 Stellar Atmosphere Programs and 2 km/s grid. (CD-ROM no. 13, Smithsonian Astrophysical Observatory, 1993).
- Alonso, A., Arribas, S. & Martínez-Roger, C. The effective temperature scale of giant stars (F0–K5). II. Empirical calibration of T_{eff} versus colours and [Fe/H]. *Astron. Astrophys.* **140**, 261–277 (1999).
- Gratton, R. G., Carretta, E. & Castelli, F. Abundances of light elements in metal-poor stars. I. Atmospheric parameters and a new T_{eff} scale. *Astron. Astrophys.* **314**, 191–203 (1996).
- Gratton, R. G., Carretta, E., Eriksson, K. & Gustafsson, B. Abundances of light elements in metal-poor stars. II. Non-LTE abundance corrections. *Astron. Astrophys.* **350**, 955–969 (1999).
- Campbell, S. W. & Lattanzio, J. C. Evolution and nucleosynthesis of extremely metal-poor and metal-free low- and intermediate-mass stars. I. Stellar yield tables and the CEMP. *Astron. Astrophys.* **490**, 769–776 (2008).
- Marigo, P. & Aringer, B. Low-temperature gas opacity. A&SOPUS: a versatile and quick computational tool. *Astron. Astrophys.* **508**, 1539–1569 (2009).
- Reimers, D. Circumstellar absorption lines and mass loss from red giants. *Mem. Soc. R. Sci. Liege* **8**, 369–382 (1975).
- Clem, J. L., VandenBerg, D. A., Grundahl, F. & Bell, R. A. Empirically constrained color-temperature relations. II. uvby. *Astron. J.* **127**, 1227–1256 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Momany of the European Southern Observatory (ESO, Chile) for providing his *UBV* photometric data set, which is mentioned in Supplementary Information section 2. S.W.C. acknowledges support from the Australian Research Council's Discovery Projects funding scheme (project DP1095368). R.J.S. is the recipient of a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation. F.G. acknowledges funding for the Stellar Astrophysics Centre provided by The Danish National Research Foundation. The research was supported by the ASTERISK project funded by the European Research Council (grant agreement no. 267864). This work was based on observations made with ESO telescopes at the La Silla Paranal Observatory under programme ID 089.D-0038 (principal investigator S.W.C.) and made extensive use of the SIMBAD, VizieR, 2MASS and NASA ADS databases.

Author Contributions S.W.C. designed and prepared the ESO Very Large Telescope (VLT) observing proposal, collected the spectroscopic data, and prepared the paper. V.D. reduced and analysed the spectroscopic data, and prepared the paper. D.Y. designed and prepared the ESO/VLT observing proposal and assisted in the paper preparation. T.N.C. calculated the stellar models and prepared figures for the paper. J.C.L. assisted in the preparation of the observing proposal and with the paper preparation. R.J.S., G.C.A. and E.C.W. assisted in the paper preparation and made preliminary observations with the Anglo-Australian Telescope. F.G. provided the *uvby* photometric data for the AGB and red giant branch sample and assisted in the paper preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.W.C. (simon.campbell@monash.edu).

The spin Hall effect in a quantum gas

M. C. Beeler^{1†}, R. A. Williams^{1†}, K. Jiménez-García^{1,2†}, L. J. LeBlanc¹, A. R. Perry¹ & I. B. Spielman¹

Electronic properties such as current flow are generally independent of the electron's spin angular momentum, an internal degree of freedom possessed by quantum particles. The spin Hall effect, first proposed 40 years ago¹, is an unusual class of phenomena in which flowing particles experience orthogonally directed, spin-dependent forces—analogue to the conventional Lorentz force that gives the Hall effect, but opposite in sign for two spin states. Spin Hall effects have been observed for electrons flowing in spin-orbit-coupled materials such as GaAs and InGaAs (refs 2, 3) and for laser light traversing dielectric junctions⁴. Here we observe the spin Hall effect in a quantum-degenerate Bose gas, and use the resulting spin-dependent Lorentz forces to realize a cold-atom spin transistor. By engineering a spatially inhomogeneous spin-orbit coupling field for our quantum gas, we explicitly introduce and measure the requisite spin-dependent Lorentz forces, finding them to be in excellent agreement with our calculations. This 'atomtronic' transistor behaves as a type of velocity-insensitive adiabatic spin selector, with potential application in devices such as magnetic⁵ or inertial⁶ sensors. In addition, such techniques for creating and measuring the spin Hall effect are clear prerequisites for engineering topological insulators^{7,8} and detecting their associated quantized spin Hall effects in quantum gases. As implemented, our system realizes a laser-actuated analogue to the archetypal semiconductor spintronic device, the Datta–Das spin transistor^{9,10}.

The spin Hall effect (SHE) is generated by spin-dependent forces orthogonal to a particle's motion—akin to the Lorentz force—that can act on electrons^{2,3,11}, photons⁴ or, as here, neutral atoms. Each of these has an internal, or 'spin', degree of freedom (a generalization of conventional quantum mechanical spin) that can be either up or down, creating a spin-1/2 (or pseudospin-1/2) system. In materials, microscopic spin-orbit coupling (SOC) induces the SHE in one of two primary ways: by means of an intrinsic mechanism driven directly by SOC¹² or by means of an extrinsic mechanism that additionally requires scattering from impurities^{1,13}. The motion of spins in systems with a SHE is strikingly similar to the motion of charges in an external magnetic field, but with equal and opposite effective Lorentz forces for each of the two spin states. Thus, just as the Lorentz force gives rise to the Hall effect for charged particles, spin-dependent Lorentz forces (SDLFs) generate SHEs.

In the Hamiltonian description of quantum mechanics, forces are described in terms of associated potentials. For example, a magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ is generated from a vector potential \mathbf{A} that enters into the Hamiltonian $\hat{H} = (\hat{\mathbf{p}} - q_0 \hat{\mathbf{A}})^2 / 2m$ with canonical momentum $\hat{\mathbf{p}}$, charge q_0 and mass m ('hats' on variables indicate quantum mechanical operators acting on continuous degrees of freedom). We engineered a vector potential that depends on an effective spin degree of freedom with opposite sign for the two effective spin states. This can create a SDLF and a SHE when the spins move perpendicularly to the resulting spin-dependent $\hat{\mathbf{B}}$.

More formally, this vector potential can be expressed as a vector of 2×2 matrices, leading to a relationship between the generalized vector potential $q_0 \hat{\mathbf{A}} \rightarrow \hat{\mathcal{A}}$ and generalized magnetic field $q_0 \hat{\mathbf{B}} \rightarrow \hat{\mathcal{B}}$ (ref. 14; 'checks' on variables indicate quantum mechanical operators acting in pseudo-spin space):

$$\hat{\mathcal{B}} = \nabla \times \hat{\mathcal{A}} - \frac{i}{\hbar} \hat{\mathcal{A}} \times \hat{\mathcal{A}} \quad (1)$$

The Heisenberg equations of motion show that $\hat{\mathcal{B}}$ is the generalized magnetic field in a spin-dependent Lorentz force law (Methods). The first term in equation (1) is analogous to the conventional magnetic field, and the second term is non-zero only when the vector components of $\hat{\mathcal{A}}$ do not all commute, that is, when $\hat{\mathcal{A}}$ is non-Abelian. The generalized Lorentz force for the two spin states can be equal and opposite, for example when $\hat{\mathcal{B}} = \mathcal{B}_0 \hat{\sigma}_3 \mathbf{e}_z$, where \mathcal{B}_0 describes the field's magnitude; $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\sigma}_3$ are the 2×2 Pauli matrices; and \mathbf{e}_x , \mathbf{e}_y and \mathbf{e}_z are the three Cartesian unit vectors.

Two different classes of vector potentials (unrelated by gauge transformations) lead to this magnetic field, one resulting from each term in equation (1). For example, in two-dimensional material systems, almost every possible form of linear SOC—combinations of linear Dresselhaus¹⁵ or Rashba¹⁶ couplings—is equivalent to a spatially uniform non-Abelian vector potential with $-i(\hat{\mathcal{A}} \times \hat{\mathcal{A}})/\hbar \propto \hat{\sigma}_3 \mathbf{e}_z$ (Methods and ref. 17). In contrast, we engineered a spin-orbit-coupled Hamiltonian with a spatially dependent, Abelian vector potential that produces $\nabla \times \hat{\mathcal{A}} \propto \hat{\sigma}_3 \mathbf{e}_z$.

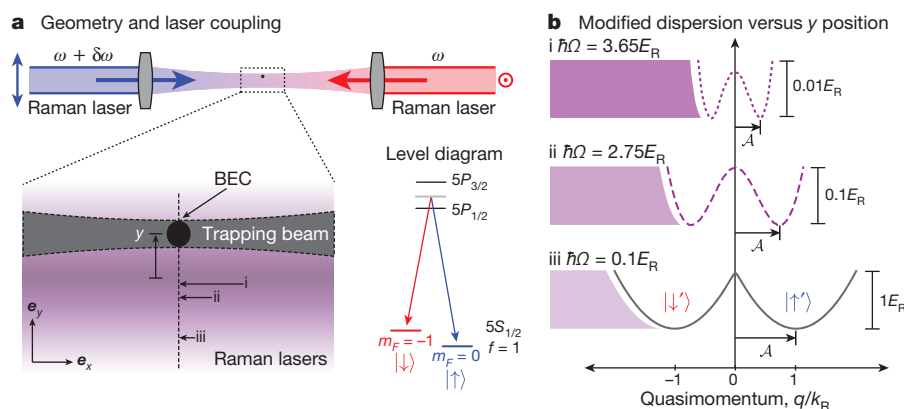


Figure 1 | Schematic of experimental set-up. **a**, Raman beams with frequencies ω and $\omega + \delta\omega$ propagating along \mathbf{e}_x coupled two states in the $f = 1$ ground-state manifold of ^{87}Rb . Dynamic control of an optical trapping beam propagating along \mathbf{e}_x allowed the BEC to be moved along \mathbf{e}_y , giving a time- and position-dependent Raman coupling. The Raman coupling altered the free-particle dispersion along \mathbf{e}_x , creating double wells²¹ in quasimomentum q . **b**, The modified dispersions, $E(q)$, shown for the three different y positions (i, ii and iii) marked in **a**. We associate states near the minimum of each well with dressed spins, and identify the location of the minima with a vector potential \mathcal{A} .

¹Joint Quantum Institute, National Institute of Standards and Technology and University of Maryland, Gaithersburg, Maryland 20899, USA. ²Departamento de Física, Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional, Mexico D.F. 07360, Mexico. [†]Present addresses: The Johns Hopkins Applied Physics Laboratory, Laurel, Maryland 20723, USA (M.C.B.); National Physical Laboratory, Teddington TW11 0LW, UK (R.A.W.); The James Franck Institute and Department of Physics, The University of Chicago, Chicago, Illinois 60637, USA (K.J.-G.).

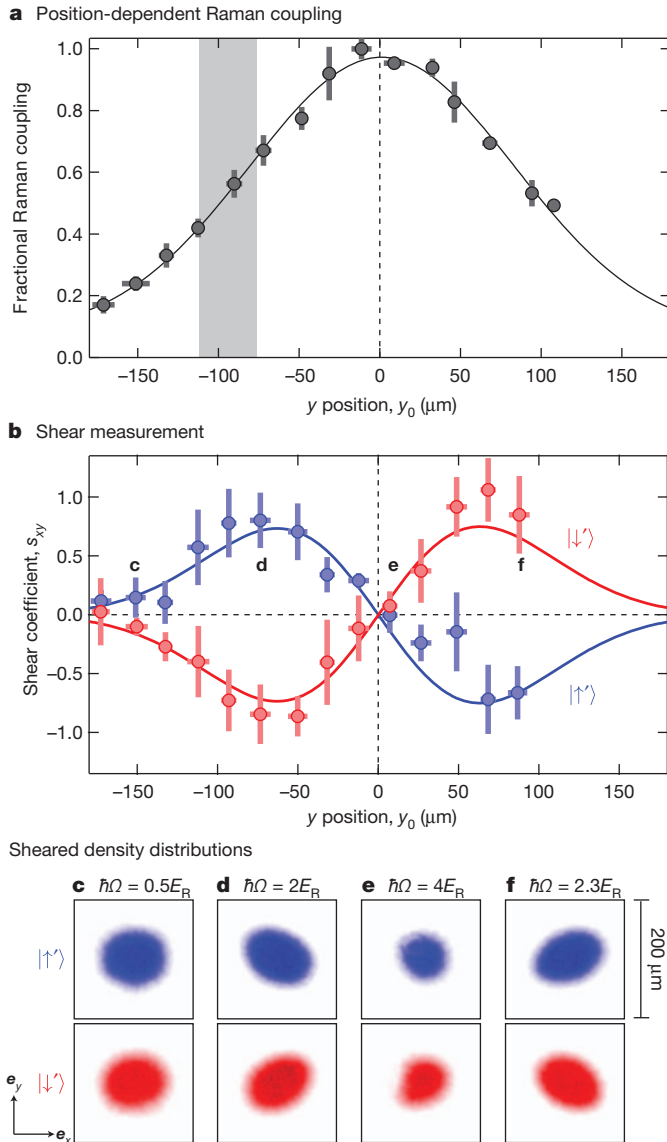


Figure 2 | Spin Hall shear. **a**, Raman coupling strength versus y position, fitted with the Raman lasers' Gaussian profile. **b**, The observed shear coefficient, s_{xy} (see text), was opposite for each spin and its magnitude followed the derivative of a Gaussian function (solid lines). All uncertainties, s.d. of ~ 5 measurements. **c–f**, Representative two-dimensional spin-momentum distributions observed after TOF at different y positions. For the data in **b–f**, the magnitude of the effect was enhanced by elongating the BEC along e_y (Methods) and so sampling a greater range of the vector potential. The measurements in Figs 3 and 4 were made in the portion of the laser shaded in grey in **a**, where $\nabla \times \mathcal{A}$ is large and nearly uniform.

The relationship between these two distinct vector potentials is unusual. Although the equations of motion describe the same SDLF leading to an intrinsic SHE, the associated energy spectra are different (for example, in the two-dimensional material systems discussed above, $[\mathcal{B}, \hat{H}] \neq 0$, implying that \mathcal{B} is time dependent in the Heisenberg picture). However, both can give rise to time-reversal-invariant topological insulators. The case with a spatially uniform vector potential mirrors the typical situation in materials where the intrinsic SOC leads to topological band structure⁸. The case with a spatially dependent vector potential leads to the simplest conceptual example of a topological insulator: two superimposed quantum Hall systems with equal but opposite magnetic fields¹⁸ (a single quantum Hall system is a topological insulator, but with broken time-reversal symmetry). Both types of vector potential exhibit the quantum SHE (QSHE) leading to topological insulators. Those resulting from spatially dependent vector potentials

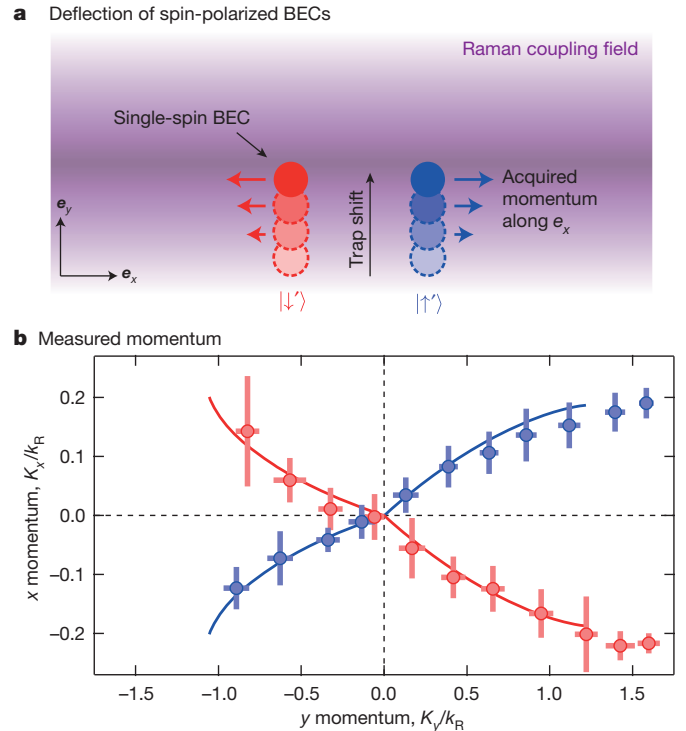


Figure 3 | Spin-polarized SHE. **a**, Spin-dependent forces along e_x from motion along e_y . **b**, Acquired momentum along e_x versus final momentum along e_y ; blue circles denote $|\uparrow\rangle$ and red circles denote $|\downarrow\rangle$. The solid curves are solutions of the Heisenberg equations of motion for each spin, fitted to the data with Ω as the only free parameter (Methods). The resulting Ω is within 15% of our measured coupling strength of $2.5(2)E_R$ at $y = -115 \mu\text{m}$, the centre of the spatial region sampled by the atoms during the measurements (grey shaded region in Fig. 2a). Uncertainties (in plotted momenta and Ω), s.d. of ~ 5 measurements.

are a direct extension of the quantum gas SHE demonstrated in this work¹⁹ but are impractical in material systems (Methods and Supplementary Information).

We realized the SHE with ultracold atoms following the proposal of ref. 20 by subjecting pseudospin-1/2 ^{87}Rb Bose-Einstein condensates (BECs) to spin- and space-dependent vector potentials. Two laser beams (which we will refer to as 'Raman lasers') with wavelength λ , propagating in opposite directions parallel to e_x , coupled the $|f=1; m_F=0, -1\rangle = |\uparrow, \downarrow\rangle$ spin states comprising our pseudospin-1/2 system (in analogy to the spin-1/2 electron) with strength Ω (Fig. 1a). The wavelength determines the single-photon recoil energy, $E_R = \hbar^2 k_R^2 / 2m$, momentum, $\hbar k_R = 2\pi\hbar/\lambda$, and velocity, $v_R = \hbar k_R / m$, where m is the mass of a ^{87}Rb atom and $2\pi\hbar$ is Planck's constant. In this configuration, the Hamiltonian describing motion along e_x includes an effective SOC term^{21–24}, altering the dispersion relation as shown in Fig. 1b. Thus modified, the dispersion relation of these laser-dressed atoms features two degenerate wells, each displaced from zero by an amount $\mathcal{A} = k_R [1 - (\hbar\Omega/4E_R)^2]^{1/2}$ for $\hbar\Omega < 4E_R$ (Methods Summary). Particles with momenta near these minima can be thought of as dressed spin states $|\uparrow', \downarrow'\rangle$ (which we will colloquially refer to as spin states when no ambiguity is possible) in the presence of a vector potential $\mathcal{A} = \mathcal{A} \hat{e}_x$. Given that Ω depends on the intensity of the Raman lasers, \mathcal{A} has the spatial dependence of the Raman lasers' Gaussian intensity profile.

The spatial dependence of \mathcal{A} gives rise to a SHE in our quantum gas^{20,25}. To probe the mechanism underlying the SHE, we abruptly changed \mathcal{A} and observed spin-dependent shearing of the atomic cloud (Fig. 2). We then observed—for a time-independent \mathcal{A} —the resulting SHE using two techniques: we propelled atoms in either state $|\uparrow'\rangle$ or state $|\downarrow'\rangle$ along e_y and detected a spin-dependent Lorentz-like response along $\pm e_x$ (Fig. 3); and, using a mixture of both dressed spin states, we used the SDLF to realize a spin transistor (Fig. 4).

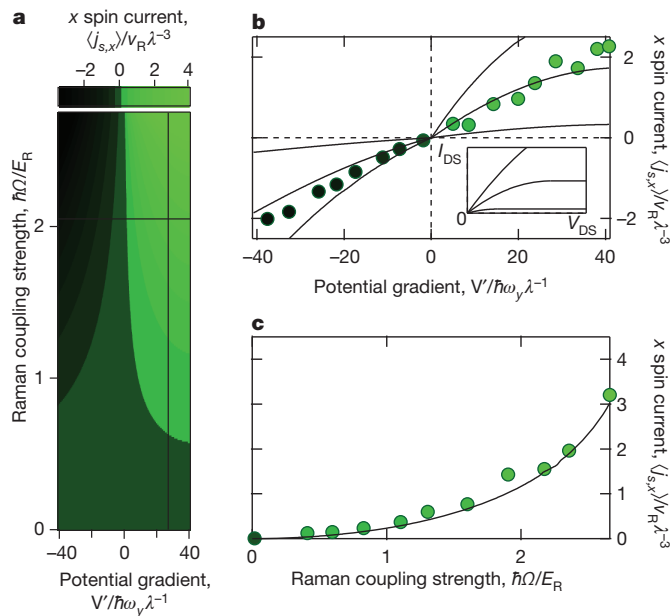


Figure 4 | Spin Hall currents. **a**, Calculated spin current versus potential gradient, V' , and coupling strength, Ω . The two cuts (black lines) show the parameters at which measurements of $\langle j_{s,x} \rangle$ were made. **b** Spin current, $\langle j_{s,x} \rangle$, versus V' . We note that $\hbar \Omega(y = -115 \mu\text{m}) = 2.3(2)E_R$. The central solid curve is a fit of our model to the data, with fitted value $\hbar \Omega_f = 2.05E_R$. The remaining curves are the modelled response at $\hbar \Omega = 1E_R$ (lower magnitude) and $2.5E_R$ (higher magnitude). Inset: FET drain-source current, I_{DS} , versus drain-source voltage, V_{DS} , for three different gate-source voltages, $V_{GS} = 1.5, 2.4$ and 3.5 V above threshold. **c**, Dependence of spin current on coupling strength at a fixed trap displacement from $y_0 = -122 \mu\text{m}$ to $y_f = -95 \mu\text{m}$ with $V' \approx 27\lambda/\hbar\omega_y$. The solid curve is the fit of our model to the data (Methods). The scatter in the data is reflective of typical uncertainties.

These experiments began with BECs of 5×10^4 atoms prepared in $|\uparrow\rangle$, $|\downarrow\rangle$ or mixtures thereof, confined in a crossed-beam optical dipole trap with typical axial frequencies $\omega_x/2\pi \approx 35$ Hz, $\omega_y/2\pi \approx 35$ Hz and $\omega_z/2\pi \approx 100$ Hz. The $\lambda = 790.13$ -nm Raman laser beams, travelling along $\pm \mathbf{e}_x$, had $170\text{-}\mu\text{m}$ waists ($1/e^2$ radius, where e is Euler's number). We moved the BECs along \mathbf{e}_y , sampling this inhomogeneous Raman laser profile, by displacing the appropriate trap beam. At any given initial y position, y_0 , we then adiabatically turned on the Raman lasers in 150 ms, Raman-dressing the BEC²¹ and transforming our initial spin states into their dressed counterparts, at rest^{26,27} (Methods).

We explored the spin and space dependence of the vector potential $\mathcal{A}(y)$ by observing the response of the BECs to abrupt temporal changes in \mathcal{A} . When \mathcal{A} depended on y , these changes sheared the BECs' density distribution. We prepared spin-polarized BECs at a variable position, y_0 (Methods). Each Raman-dressed BEC therefore sampled a range of Raman coupling strengths across its $40\text{-}\mu\text{m}$ diameter (Fig. 2a). When the Raman lasers were abruptly turned off, the initially motionless BEC experienced a spin-dependent 'electric' force, $-\partial \mathcal{A}/\partial t$, resulting from a time-changing vector potential along \mathbf{e}_x (ref. 28). We probed this system by switching off the dipole trap and the Raman beams in less than $1 \mu\text{s}$ and absorption-imaging the atoms after a 30-ms time of flight (TOF; its duration was common to all of our measurements).

Because $\mathcal{A}(y)$ depended on both spin and y position, we observed a spin- and y_0 -dependent shear²⁹ in the density distribution (Fig. 2b) after TOF, described by $n(x, y, z) \propto 1 - (x/R_x)^2 - (y/R_y)^2 - (z/R_z)^2 - s_{xy}xy/R_xR_y$, where R_x , R_y and R_z are the BEC's Thomas-Fermi radii. The spatial dependence of \mathcal{A} is quantified by the shear coefficient, s_{xy} , which is obtained by fitting this distribution (integrated along \mathbf{e}_z) to the TOF BEC density distribution. The spin-dependent nature of the vector potential is evident in the opposite sign of the shear for each spin

(Fig. 2b–f) and in that the magnitude of the shear coefficient follows the local derivative of the vector potential at the centre of the BEC.

We first observed the SHE using spin-polarized BECs. This would be atypical in condensed-matter systems, where both spins are usually present. After preparing a spin-polarized BEC at a position, y_0 , between $y_{\min} = -135 \mu\text{m}$ and $y_{\max} = -95 \mu\text{m}$ (grey shaded region in Fig. 2a, a region over which the SDLF was both reasonably large and uniform), we abruptly displaced the centre of the harmonic trap to $y_f = y_{\max}$ or $y_f = y_{\min}$. This displacement can formally be understood as resulting from an applied potential with gradient V' . The atoms accelerated to a final y momentum $\hbar K_y$ in ~ 7 ms (one-quarter of the \mathbf{e}_y -trap period). During this time, the SDLF accelerated the atoms perpendicular to their instantaneous momentum, resulting in a final x -momentum $\hbar K_x$. By waiting this quarter-period after trap displacement, we ensured that the atoms always arrived at y_f (regardless of the choice of y_0). Subsequently, the trap was turned off abruptly (in less than $1 \mu\text{s}$), the Raman lasers were turned off slowly by comparison with dressed-state bandgaps ($\sim 500 \mu\text{s}$), and the atoms were imaged after TOF to determine their final momentum (Fig. 3b). With this turn-off procedure, the atoms experienced a force $-\partial \mathcal{A}/\partial t$ along \mathbf{e}_x (independent of y_0) that shifted the final centre-of-mass position after TOF from the observed position of atoms released in the absence of \mathcal{A} (Methods). We calibrated this zero-momentum TOF position by detecting atoms released from rest at y_f .

Each spin-polarized BEC acquired a momentum along \mathbf{e}_x that was directed oppositely for the two spins and related to its final momentum along \mathbf{e}_y , demonstrating an intrinsic SHE. We modelled the dynamics of each spin (Fig. 3b, solid curves, and Methods) by solving the Heisenberg equations of motion. Because our atoms remain in the lowest-energy band plotted in Fig. 1b, the Heisenberg equations of motion reduce to classical dynamics subject to the spin-orbit-coupled dispersion shown in Fig. 1b. The model predicts both K_x and K_y as functions of initial and final trap displacement. We leave Ω as a fitting parameter, the value of which is within 15% of our calibrated value. The results of this model are plotted along with the data in Fig. 3b.

Next we realized the SHE in a configuration analogous to solid systems by using mixtures of both spins. In the presence of both spins, we define average spin and particle current densities $\langle \mathbf{j}_s \rangle = \langle \mathbf{j}_{\uparrow} \rangle - \langle \mathbf{j}_{\downarrow} \rangle$ and $\langle \mathbf{j}_p \rangle = \langle \mathbf{j}_{\uparrow} \rangle + \langle \mathbf{j}_{\downarrow} \rangle$, where the average current density for spin i (either \uparrow or \downarrow) is $\langle \mathbf{j}_i \rangle = (1/V) \int_V n_i(\mathbf{r}) \mathbf{v}_i(\mathbf{r}) d\mathbf{r}$, with density n , velocity \mathbf{v} , and *in situ* BEC volume V . An equal current of each spin moving in the same direction corresponds to a pure particle current, and an equal current of each spin moving in opposite directions gives a pure spin current.

This third class of experiments started with BECs in a mixture of both spins (Methods). We generated a pure particle current using the trap displacement technique described above. As before, the system evolved under the SDLF for ~ 7 ms, after which time the atoms were released from the trap and the Raman lasers adiabatically turned off (Methods). Each TOF image contained information about both dressed spin states, allowing us to determine the spin and particle currents simultaneously. We modelled the resulting spin current along \mathbf{e}_x as a function of coupling strength, Ω , and potential gradient, V' . The system's spin response, $\langle j_{s,x} \rangle = \langle \mathbf{j}_s \rangle \cdot \mathbf{e}_x$, is shown in Fig. 4a. By varying one parameter at a time (Fig. 4a, black lines), we measured the spin current as a function of V' (Fig. 4b) and as a function of Ω (Fig. 4c). In both cases, the experiment agrees with our model.

Despite the existence of the SHE in all spin-orbit-coupled metals and semiconductors, the technology for studying the SHE was developed only recently. Soon afterwards, the SHE was exploited to develop spintronic devices¹¹. In this spirit, our experiment describes an externally actuated 'atomtronic' bipolar spin transistor^{9,10}, where Ω plays the part of the transistor's gate voltage and the potential gradient, V' , is analogous to the drain-source voltage. The spin current turns on abruptly at $\hbar \Omega \approx 1E_R$ (Fig. 4c), with a final spin current set by the potential gradient. For a given Raman coupling ('gate voltage'),

however, the spin current turns on smoothly with positive or negative potential gradient (Fig. 4b). This similarity between our system and a field-effect transistor (FET) is further highlighted in Fig. 4b, where the three black curves modelling our system's response at three different Raman coupling strengths are compared with the characteristic response of a FET's drain-source current as a function of drain-source voltage at three different gate-source voltages.

In atomic systems, other techniques can separate particles according to spin, such as the well-known Stern–Gerlach effect. Our technique complements these, because the spin-dependent force depends not on the atoms' positions (as in the Stern–Gerlach effect), but on their velocities. For example, a device with a finite SHE interaction region will deflect an incoming atomic beam by an amount independent of the velocity with which the atoms enter the region; although an increase in initial velocity decreases the interaction time, the perpendicular force increases (for interaction times much less than $2\pi m/B_0$). For devices using the Stern–Gerlach effect, the deflection depends only on the interaction time, which changes with initial velocity. A spin transistor might operate using either our SDF or a Stern–Gerlach-type force, but in each case its behaviour will be quite different. For example, using our transistor as the input and output beam splitter in Mach–Zehnder-type inertial sensors⁵ could yield coherent adiabatic momentum splitting that is independent of the atoms' longitudinal velocity profile.

We have demonstrated an intrinsic SHE in a quantum gas using a precisely engineered spin- and space-dependent vector potential. Systems such as this—with the experimental parameters available at present—are candidates for a.c. gravity gradiometers, when applied to dilute clouds where interaction effects are negligible⁶. In addition, time-reversal-invariant topological insulators manifest the QSHE¹⁸. Using present technologies, our method for producing the SHE could produce the QSHE in an ultracold gas of fermionic ⁴⁰K (Methods and Supplementary Information). Despite the technical challenges, the simplicity of our setup—two atomic spin states and two oppositely directed lasers—makes our approach an appealing method for achieving the QSHE. In similar parameter regimes, it may be possible to realize exotic, interacting topological insulators using Bose gases^{19,30}.

METHODS SUMMARY

System preparation. A bias magnetic field of $B_0 = 2.1$ mT lifted the degeneracy of the $|f = 1, m_F = 0, \pm 1\rangle$ spin states in the electronic ground-state manifold of ⁸⁷Rb, leading to an energy-level splitting, $\Delta E = 2\pi\hbar \times 15$ MHz, between $|m_F = -1\rangle$ and $|m_F = 0\rangle$ that matched the $\hbar\delta\omega$ energy difference between the Raman laser beams' photons, where $\delta\omega$ is the frequency difference between the Raman beams. Owing to the large bias field, the $|m_F = +1\rangle$ spin state was detuned from Raman resonance by $17.8E_R$ and was inactive in our experiments.

In the limit of zero Raman coupling, each dressed spin continuously connects to a bare spin with quasimomentum $\hbar|q| = \hbar k_R$. To load a specific dressed spin, we started with a BEC in $|m_F = -1\rangle$, $|m_F = 0\rangle$ or a mixture of those states and turned on the Raman lasers in 150 ms. During experiments on spin-polarized BECs, we avoided any undesired population of the other dressed spin state by applying a detuning $\hbar\delta = \Delta E - \hbar\delta\omega = 0.15E_R$ during the ramp up of Ω and then by shifting to resonance ($\delta = 0$) with a 1-ms ramp of B_0 . An acousto-optic modulator shifted the position of the dipole trap beam propagating along \mathbf{e}_y , allowing controlled translation of the atomic sample along \mathbf{e}_y .

Dressed states. The single particle properties of our system are well described by the Hamiltonian²¹

$$\hat{H} = \frac{\hbar^2(\hat{q}_x^2 + \hat{q}_y^2 + \hat{q}_z^2)}{2m} \hat{\mathbb{I}} + \frac{\hbar\Omega}{2} \hat{\sigma}_1 - \frac{\hbar^2 k_R \hat{q}_y}{m} \hat{\sigma}_3 + E_R \hat{\mathbb{I}}$$

for resonant Raman coupling, as we use here. The eigenenergies

$$\mathcal{E}_{\pm}(q) + \frac{\hbar^2(k_y^2 + k_z^2)}{2m}$$

with

$$\mathcal{E}_{\pm}(q) = \frac{\hbar^2 q^2}{2m} + E_R \pm \sqrt{\left(\frac{\hbar\Omega}{2}\right)^2 + \left(\frac{\hbar^2 k_R q}{m}\right)^2}$$

define a pair of effective dispersion relations, the lower of which, $\mathcal{E}_{-}(q)$, is plotted for $k_y = k_z = 0$ in Fig. 1b for a selection of coupling strengths.

Full Methods and any associated references are available in the online version of the paper.

Received 12 December 2012; accepted 9 April 2013.

Published online 5 June 2013.

1. Dyakonov, M. I. & Perel, V. I. Possibility of orienting electron spins with current. *Sov. Phys. JETP* **13**, 467–469 (1971).
2. Kato, Y. K., Myers, R. C., Gossard, A. C. & Awschalom, D. D. Observation of the spin Hall effect in semiconductors. *Science* **306**, 1910–1913 (2004).
3. Wunderlich, J., Kaestner, B., Sinova, J. & Jungwirth, T. Experimental observation of the spin-Hall effect in a two-dimensional spin-orbit coupled semiconductor system. *Phys. Rev. Lett.* **94**, 047204 (2005).
4. Hosten, O. & Kwiat, P. Observation of the spin Hall effect of light via weak measurements. *Science* **319**, 787–790 (2008).
5. Kitching, J., Knappe, S. & Donley, E. Atomic sensors – a review. *Sensors J. IEEE* **11**, 1749–1758 (2011).
6. Anderson, B. M., Taylor, J. M. & Galitski, V. M. Interferometry with synthetic gauge fields. *Phys. Rev. A* **83**, 031602 (2011).
7. Kane, C. L. & Mele, E. J. Z_2 topological order and the quantum spin Hall effect. *Phys. Rev. Lett.* **95**, 146802 (2005).
8. Bernevig, B. A., Hughes, T. L. & Zhang, S.-C. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314**, 1757–1761 (2006).
9. Vaishnav, J. Y., Ruseckas, J., Clark, C. W. & Juzeliūnas, G. Spin field effect transistors with ultracold atoms. *Phys. Rev. Lett.* **101**, 265302 (2008).
10. Datta, S. & Das, B. Electronic analog of the electro-optic modulator. *Appl. Phys. Lett.* **56**, 665–667 (1990).
11. Jungwirth, T., Wunderlich, J. & Olejnik, K. Spin Hall effect devices. *Nature Mater.* **11**, 382–390 (2012).
12. Sinova, J. *et al.* Universal intrinsic spin Hall effect. *Phys. Rev. Lett.* **92**, 126603 (2004).
13. Hirsch, J. E. Spin Hall effect. *Phys. Rev. Lett.* **83**, 1834–1837 (1999).
14. Ruseckas, J., Juzeliūnas, G., Öhberg, P. & Fleischhauer, M. Non-Abelian gauge potentials for ultracold atoms with degenerate dark states. *Phys. Rev. Lett.* **95**, 010404 (2005).
15. Dresselhaus, G. Spin-orbit coupling effects in zinc blende structures. *Phys. Rev.* **100**, 580–586 (1955).
16. Bychkov, Y. A. & Rashba, E. I. Oscillatory effects and the magnetic susceptibility of carriers in inversion layers. *J. Phys. C* **17**, 6039 (1984).
17. Lyanda-Geller, Y. Quantum interference and electron-electron interactions at strong spin-orbit coupling in disordered systems. *Phys. Rev. Lett.* **80**, 4273–4276 (1998).
18. Hasan, M. Z. & Kane, C. L. Topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
19. Stanescu, T. D., Galitski, V., Vaishnav, J. Y., Clark, C. W. & Das Sarma, S. Topological insulators and metals in atomic optical lattices. *Phys. Rev. A* **79**, 053639 (2009).
20. Zhu, S.-L., Fu, H., Wu, C.-J., Zhang, S.-C. & Duan, L.-M. Spin Hall effects for cold atoms in a light-induced gauge potential. *Phys. Rev. Lett.* **97**, 240401 (2006).
21. Lin, Y. J., Jiménez-García, K. & Spielman, I. B. Spin-orbit-coupled Bose–Einstein condensates. *Nature* **471**, 83–86 (2011).
22. Zhang, J.-Y. *et al.* Collective dipole oscillations of a spin-orbit coupled Bose–Einstein condensate. *Phys. Rev. Lett.* **109**, 115301 (2012).
23. Wang, P. *et al.* Spin-orbit coupled degenerate Fermi gases. *Phys. Rev. Lett.* **109**, 095301 (2012).
24. Cheuk, L. W. *et al.* Spin-injection spectroscopy of a spin-orbit coupled Fermi gas. *Phys. Rev. Lett.* **109**, 095302 (2012).
25. Liu, X.-J., Liu, X., Kwek, L. C. & Oh, C. H. Optically induced spin-Hall effect in atoms. *Phys. Rev. Lett.* **98**, 026602 (2007).
26. Zhang, Y., Mao, L. & Zhang, C. Mean-field dynamics of spin-orbit coupled Bose–Einstein condensates. *Phys. Rev. Lett.* **108**, 035302 (2012).
27. Ho, T.-L. & Zhang, S. Bose–Einstein condensates with spin-orbit interaction. *Phys. Rev. Lett.* **107**, 150403 (2011).
28. Lin, Y.-J. *et al.* A synthetic electric force acting on neutral atoms. *Nature Phys.* **7**, 531–534 (2011).
29. LeBlanc, L. J. *et al.* Observation of a superfluid Hall effect. *Proc. Natl Acad. Sci. USA* **109**, 10811–10814 (2012).
30. Dalibard, J., Gerbier, F., Juzeliūnas, G. & Öhberg, P. Artificial gauge potentials for neutral atoms. *Rev. Mod. Phys.* **83**, 1523–1543 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was partially supported by the DARPA OLE programme; the ARO atomtronics MURI, NIST, and the US NSF through the PFC at the JQI. M.C.B. acknowledges NIST-ARRA, L.J.L. acknowledges support from NSERC and K.J.-G. acknowledges CONACYT.

Author Contributions M.C.B. led the data-taking effort, in which all co-authors participated. M.C.B. carried out the analysis, M.C.B. and I.B.S. performed theoretical and analytical calculations, and all authors contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.B.S. (ian.spielman@nist.gov).

METHODS

System preparation. A $B_0 = 2.1$ -mT bias magnetic field lifted the degeneracy of the $|f = 1, m_F = 0, \pm 1\rangle$ spin states in the electronic ground-state manifold of ^{87}Rb , leading to an energy-level splitting, $\Delta E = 2\pi\hbar \times 15$ MHz, between $|m_F = -1\rangle$ and $|m_F = 0\rangle$ that matched the $\hbar\delta\omega$ energy difference between the Raman laser beams' photons, where $\delta\omega$ is the frequency difference between the Raman beams. Owing to the large bias field, the $|m_F = +1\rangle$ spin state was detuned from Raman resonance by $17.8E_R$ and was inactive in our experiments.

In the limit of zero Raman coupling, each dressed spin continuously connects to a bare spin with quasimomentum $\hbar|q| = \hbar k_R$. To load a specific dressed spin, we started with a BEC in $|m_F = -1\rangle$, $|m_F = 0\rangle$ or a mixture of those states, and turned on the Raman lasers in 150 ms. During experiments on spin-polarized BECs, we avoided any undesired population of the other dressed spin by applying a detuning $\hbar\delta = \Delta E - \hbar\delta\omega = 0.15E_R$ during the ramp up of Ω then by shifting to resonance ($\delta = 0$) with a 1-ms ramp of B_0 . An acousto-optic modulator shifted the position of the dipole trap beam propagating along \mathbf{e}_x , allowing controlled translation of the atomic sample along \mathbf{e}_y .

Dressed states. The single particle properties of our system are well-described by the Hamiltonian²¹

$$\hat{H} = \frac{\hbar^2(\hat{q}^2 + \hat{k}_y^2 + \hat{k}_z^2)}{2m} \hat{I} + \frac{\hbar\Omega}{2} \hat{\sigma}_1 - \frac{\hbar^2 k_R \hat{q}}{m} \hat{\sigma}_3 + E_R \hat{I}$$

for resonant Raman coupling, as we use here. The eigenenergies

$$\mathcal{E}_{\pm}(q) + \frac{\hbar^2(k_y^2 + k_z^2)}{2m}$$

with

$$\mathcal{E}_{\pm}(q) = \frac{\hbar^2 q^2}{2m} + E_R \pm \sqrt{\left(\frac{\hbar\Omega}{2}\right)^2 + \left(\frac{\hbar^2 k_R q}{m}\right)^2}$$

define a pair of effective dispersion relations, the lower of which, $\mathcal{E}_{-}(q)$, is plotted for $k_y = k_z = 0$ in Fig. 1b for a selection of coupling strengths.

Quantum SHE. Our technique for producing the SHE can be extended to realize the QSHE in two-dimensional ultracold Fermi gases. A simple example system that exhibits the QSHE can be constructed by overlapping two integer quantum Hall systems with filling factors of $\nu = 1$ and opposite magnetic field, the second of which implies that they have opposite Chern numbers¹⁸. Although this construct—two separate, spatially overlapped electron systems that experience opposite magnetic field—is artificial, the QSHE can arise from an equal mixture of spins experiencing strong opposite, spin-dependent 'magnetic' fields.

To understand intuitively how this might work, we consider our effective pseudospin Hamiltonian in two dimensions for $\hbar\Omega < 4E_R$ (ignoring the optical confinement, the scalar light shift from the Raman lasers and the zero-energy shift from the Raman dressing):

$$\hat{H} = \frac{1}{2m^*} (\hat{\mathbf{p}} \hat{I} - \mathcal{A} \hat{\sigma}_3 \mathbf{e}_x)^2$$

Here \hat{I} is the 2×2 identity matrix, $\mathcal{A} = \hbar k_R [1 - (\hbar\Omega/4E_R)^2]^{1/2}$ is the magnitude of the Raman-laser-induced vector potential, $\hat{\mathbf{A}} = \mathcal{A} \hat{\sigma}_3 \mathbf{e}_x$ is the matrix-valued vector potential, $\hat{\mathbf{p}}$ is the canonical momentum and m^* is the effective mass tensor. Here, pseudospin is a good quantum number and the system can be thought of as two independent systems that respond oppositely to temporal and spatial gradients in \mathcal{A} . By introducing a large, non-zero $\nabla \times \hat{\mathbf{A}}$, each spin state taken separately could be driven to the integer quantum Hall regime, thereby creating a QSHE in a system composed of an equal mixture of both spins.

Our specific proposal to extend our work and realize the QSHE uses ^{40}K confined in a quasi-two-dimensional geometry in the \mathbf{e}_x - \mathbf{e}_y plane. Two Raman lasers counterpropagating along \mathbf{e}_x couple together two magnetic sublevels in the $|f = 9/2\rangle$ ground-state manifold. Tailoring the Raman lasers (using a spatial light modulator^{21–24}, for instance) to have a position-dependent coupling $\hbar\Omega(y) = 4E_R \sqrt{L_y^2 - y^2} / L_y$ for $y \in (0, L_y]$ along \mathbf{e}_y produces a linearly varying \mathcal{A} . Each pseudospin experiences an oppositely directed, uniform, synthetic magnetic field with cyclotron frequency $\omega_c = \hbar k_R / m L_y$ for $y \in (0, L_y]$.

To reach the QSHE regime, the thermal energy scale, $k_B T$, the Fermi energy, \mathcal{E}_F , and the cyclotron energy, $\hbar\omega_c$, must satisfy $k_B T < \mathcal{E}_F \approx \hbar\omega_c$ (so that the Fermi energy falls in the gap between the ground and first Landau levels). Here, k_B is Boltzmann's constant and T is the temperature. The cyclotron frequency therefore sets the energy scales necessary to produce a QSHE. For realistic system sizes of 5–10 μm , the cyclotron frequency is $\omega_c/2\pi \approx 100$ Hz. In Supplementary Information, we make this argument rigorous for our actual experimental configuration.

Notes on Fig. 2. The Raman coupling strength in Fig. 2a was measured as described in refs 23, 35. For the data in Fig. 2b–f, the aspect ratio of the BEC was adjusted from its typical cylindrical symmetry to be 50% longer along \mathbf{e}_y than along \mathbf{e}_x by adjusting the optical trap, and the atom number was maintained at $>10^5$.

Measurement and analysis. To measure the atoms' momenta, the optical confinement was turned off abruptly while the Raman lasers' intensity was linearly ramped to zero in 0.5–1 ms. This procedure transferred each dressed spin to a bare spin moving with an x momentum equal to its quasimomentum q and a y momentum equal to its in-trap y momentum, K_y . A magnetic field gradient applied for a few milliseconds during the 30-ms TOF separated the two bare spins along \mathbf{e}_y through the Stern–Gerlach effect. After this separation, we measured the atomic density distribution and obtained its mean position. To determine the atoms' *in situ* momenta, we referenced the measured TOF positions to the TOF positions observed for atoms under the same experimental conditions, but at rest. For example, when the trap was suddenly displaced as in Fig. 3 or 4, the reference position was determined by adiabatically dressing the atoms at the final trap position and measuring the TOF position. Subtracting the TOF position of the abruptly displaced atoms from the reference TOF position allowed us to determine the in-trap momentum.

This measurement of the momenta contained two contributions that biased the TOF positions away from the actual momenta. If the atoms do not reach their equilibrium position in the trap before TOF begins, our subtraction procedure does not yield the actual velocity, because this initial displacement is interpreted as momentum after TOF. According to our simulations, this resulted in a systematic underestimation of the momentum along \mathbf{e}_x and \mathbf{e}_y . In addition, to compensate gravity during displacement of the optical trap, the overall intensity of the optical trapping beams was increased by 25% at the same time as the position of the optical trap was changed. Owing to the competition between the optical trap and the near-linear spatial dependence of the energy minimum of the Raman-dressed bands, this power increase shifted the equilibrium position of the atoms along \mathbf{e}_y even in the absence of an optical trap displacement. We measured the equilibrium position of our atoms by increasing the power of the optical trap for ~ 7 ms without displacing it, leading to a small difference in our measured zero momentum from the actual zero momentum. These effects, which result in a momentum correction of up to 20%, were all included in our simulations.

Small fluctuations in our laboratory magnetic bias field lift the energy degeneracy of the two pseudospin states, leading to fluctuations in the pseudospin population distribution. When working with a mixture of pseudospins, we discarded any measurement for which the population of one spin state was greater than 150% of the other, resulting in up to 60% of the data from each sequence being omitted from analysis. In addition, when both dressed spins were used together, there was an initial spatial segregation of the spins owing to a repulsive interaction between them^{21,26,27}. Although the *in situ* spatial distribution of the spins was modified before the experiment began, this interaction energy did not significantly affect our momentum measurements, because the *in situ* displacement was small compared with the typical TOF displacements giving the momentum signal.

Simulations. Because transitions between the dressed-spin bands are energetically suppressed owing to the large energy gap between bands (compared with the energy of the dynamics), the Heisenberg equations of motion for our system were the same as Hamilton's classical equations of motion in the lowest band. In our simulation, the classical Hamiltonian included the modified position-dependent dispersion relation along \mathbf{e}_x (Fig. 1b), the scalar potential from the Raman beams, the scalar potential from the optical dipole trap and the gravitational potential. The dispersion relation was calculated by diagonalizing our system's spin-orbit-coupled Hamiltonian²¹ and retaining only the lowest energy band. It is the position-dependent modified dispersion relation that drives the observed SHE. The solutions to Hamilton's coupled differential equations yielded values for the position and momentum (or quasimomentum) in all three spatial directions as functions of time. For a given dressed spin, the simulated mechanical momentum K_x was the difference between $q(t)$ and the location of the minimum of the dispersion curve associated with that dressed spin. Our model does not predict values of $K_y > 1.2\hbar k_R$ that were observed in the experiment, but this can be explained by deviations of our optical trap from the ideal Gaussian beams used in our model.

Linear Dresselhaus and Rashba SOC as a vector potential. Consider the Rashba and linear Dresselhaus SOC Hamiltonians in two dimensions³⁶:

$$\tilde{\mathcal{H}}_R = \frac{\alpha}{\hbar} (\hat{p}_x \hat{\sigma}_2 - \hat{p}_y \hat{\sigma}_1)$$

and

$$\tilde{\mathcal{H}}_D = \frac{\beta}{\hbar} (\hat{p}_y \hat{\sigma}_2 - \hat{p}_x \hat{\sigma}_1)$$

Here \hat{p}_i is the momentum along the $i \in \{e_x, e_y, e_z\}$ spatial direction and α and β are the respective strengths of the Rashba and Dresselhaus SOCs. The total Hamiltonian containing both of these terms

$$\mathcal{H}_{\text{SOC}} = \frac{\hat{p}^2}{2m} + \mathcal{H}_D + \mathcal{H}_R$$

can be expressed as

$$\mathcal{H}_{\text{SOC}} = \frac{1}{2m} (\hat{p} - \mathcal{A})^2 - \frac{m}{\hbar^2} \tilde{I}(\alpha^2 + \beta^2)$$

with

$$\mathcal{A} = -\frac{m}{\hbar} (\alpha \tilde{\sigma}_2 - \beta \tilde{\sigma}_1, \beta \tilde{\sigma}_2 - \alpha \tilde{\sigma}_1, 0) \quad (2)$$

The generalized magnetic field from this vector potential is

$$\tilde{\mathcal{B}} = \nabla \times \mathcal{A} - \frac{i}{\hbar} \mathcal{A} \times \mathcal{A} = \left[\frac{2m^2}{\hbar^3} (\alpha^2 - \beta^2) \tilde{\sigma}_3 \right] e_z \quad (3)$$

Lorentz force. A generalized magnetic field defined by equation (3) gives a generalized Lorentz force law. Following ref. 37, we start with a Hamiltonian

$$\hat{H} = \frac{1}{2m} (\hat{p} - \mathcal{A})^2$$

containing a non-Abelian vector potential in three spatial dimensions with a finite number of internal degrees of freedom. The Heisenberg equation of motion for the position \hat{x} is

$$\frac{d\hat{x}_i}{dt} = \frac{1}{i\hbar} [\hat{x}_i, \hat{H}] = \frac{1}{m} (\hat{p}_i - \mathcal{A}_i) \equiv \frac{1}{m} \tilde{I}_i$$

We identify \tilde{I} as the particle's mechanical momentum. The commutator $[\tilde{I}_i, \tilde{I}_j] = i\hbar \epsilon_{ijk} \tilde{\mathcal{B}}_k$, or $\tilde{\mathcal{B}} = \nabla \times \mathcal{A} - (i/\hbar) \mathcal{A} \times \mathcal{A}$, defines the generalized magnetic field (ϵ_{ijk} is the Levi-Civita symbol). For Abelian vector potentials, the different components of \mathcal{A} all commute, and this expression for $\tilde{\mathcal{B}}$ reduces to the familiar $\mathcal{B} = \nabla \times \mathcal{A}$.

We derive the Lorentz force law starting with the Heisenberg equation of motion for the mechanical momentum:

$$\frac{d\tilde{I}}{dt} = m \frac{d^2 \hat{x}}{dt^2} = \frac{1}{i\hbar} [\tilde{I}, \hat{H}]$$

For an individual component

$$[\tilde{I}_i, \hat{H}] = \frac{i\hbar}{2m} (\tilde{\mathcal{B}}_k \tilde{I}_j - \tilde{\mathcal{B}}_j \tilde{I}_k + \tilde{I}_j \tilde{\mathcal{B}}_k - \tilde{I}_k \tilde{\mathcal{B}}_j)$$

which is the i th component of the symmetrized Lorentz force law

$$F = m \frac{d^2 \hat{x}}{dt^2} = \frac{1}{2} \left(\frac{d\hat{x}}{dt} \times \tilde{\mathcal{B}} - \tilde{\mathcal{B}} \times \frac{d\hat{x}}{dt} \right)$$

Because the $\tilde{\mathcal{B}}$ field from linear combinations of Rashba and Dresselhaus SOCs (equation (3)) and the $\tilde{\mathcal{B}}$ from our experiment are both proportional to $\tilde{\sigma}_3$, the equations of motion for the mechanical momentum in the two cases are the same. However, for the vector potential in equation (2), $\tilde{\mathcal{B}}$ does not commute with the Hamiltonian, leading to an additional Heisenberg equation of motion for $\tilde{\mathcal{B}}$ which must be included. Despite this additional complexity, the SDLF generates the SHE in both situations.

Gauge invariance. The magnetic field defined by equation (3) is not gauge invariant. The definition of gauge transformations is generalized in any discussion of non-Abelian vector potentials. For the SU(2) symmetry group, a gauge transform is a position-dependent unitary rotation in spin space^{38,39}

$$\psi \rightarrow \tilde{V}(\hat{x})\psi$$

with

$$\tilde{V}(\hat{x}) = \exp[i\alpha(\hat{x}) \cdot \tilde{\sigma}]$$

where α is an arbitrary vector of functions of \hat{x} and $\tilde{\sigma}$ is the vector of 2×2 Pauli matrices including the identity. Under this gauge transformation, the Lagrangian must remain unchanged, requiring the magnetic field to transform according to⁴⁰

$$\tilde{\mathcal{B}} \rightarrow \tilde{V}(\hat{x}) \tilde{\mathcal{B}} \tilde{V}^\dagger(\hat{x})$$

Despite the lack of gauge invariance of the magnetic field, an Abelian magnetic field cannot be gauge transformed to a non-Abelian field.

This definition for gauge transforms can be generalized to a gauge with generators from any continuous symmetry group by replacing $\tilde{\sigma}$ with a vector of the generators of the symmetry group. For instance, in the case of a scalar vector potential from classical electrodynamics with U(1) symmetry, the generator of the symmetry group is a scalar, and the gauge transformation becomes the familiar position-dependent phase.

- Palima, D., Alonzo, C. A., Rodrigo, P. J. & Glückstad, J. Generalized phase contrast matched to Gaussian illumination. *Opt. Express* **15**, 11971–11977 (2007).
- Pasienski, M. & DeMarco, B. A high-accuracy algorithm for designing arbitrary holographic atom traps. *Opt. Express* **16**, 2176–2190 (2008).
- Gaunt, A. L. & Hadzibabic, Z. Robust digital holography for ultracold atom trapping. *Sci. Rep.* **2**, 721 (2012).
- Lee, J. G., McIlvain, B. J., Lobb, C. J. & Hill, W. T. III. Analogs of basic electronic circuit elements in a free-space atom chip. *Sci. Rep.* **3**, 1034 (2013).
- Lin, Y.-J. *et al.* Bose-Einstein condensate in a uniform light-induced vector potential. *Phys. Rev. Lett.* **102**, 130401 (2009).
- Schliemann, J. Spin Hall effect. *Int. J. Mod. Phys. B* **20**, 1015–1036 (2006).
- Sakurai, J. J. *Modern Quantum Mechanics* 130–131 (Addison-Wesley, 1994).
- Peskin, M. E. & Schroeder, D. V. *An Introduction to Quantum Field Theory* 481–502 (Westview, 1995).
- Yang, C. N. & Mills, R. L. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.* **96**, 191–195 (1954).
- Estienne, B., Haaker, S. M. & Schoutens, K. Particles in non-Abelian gauge potentials: Landau problem and insertion of non-Abelian flux. *N. J. Phys.* **13**, 045012 (2011).

A temporal cloak at telecommunication data rate

Joseph M. Lukens¹, Daniel E. Leaird¹ & Andrew M. Weiner¹

Through advances in metamaterials—artificially engineered media with exotic properties, including negative refractive index^{1–3}—the once fanciful invisibility cloak has now assumed a prominent place in scientific research^{4–13}. By extending these concepts to the temporal domain¹⁴, investigators have recently described a cloak which hides events in time by creating a temporal gap in a probe beam that is subsequently closed up; any interaction which takes place during this hole in time is not detected¹⁵. However, these results are limited to isolated events that fill a tiny portion of the temporal period, giving a fractional cloaking window of only about 10^{-4} per cent at a repetition rate of 41 kilohertz (ref. 15)—which is much too low for applications such as optical communications. Here we demonstrate another technique for temporal cloaking, which operates at telecommunication data rates and, by exploiting temporal self-imaging through the Talbot effect, hides optical data from a receiver. We succeed in cloaking 46 per cent of the entire time axis and conceal pseudorandom digital data at a rate of 12.7 gigabits per second. This potential to cloak real-world messages introduces temporal cloaking into the sphere of practical application, with immediate ramifications in secure communications.

As in the first demonstration of a ‘time cloak’¹⁵, the theoretical foundation for our cloak is space-time duality, the formal mathematical equivalence between paraxial diffraction and narrowband dispersion^{16,17}. This correspondence permits the extension of concepts typically associated with spatial Fourier imaging into the time domain. For example, just as a traditional thin lens applies a quadratic phase in space, a temporal lens can be constructed that applies a quadratic phase profile in time. But although time lenses with extremely large chirp coefficients can be obtained through parametric nonlinear interactions^{18,19}, such schemes are not easily implemented at gigahertz rates. Instead, electro-optic phase modulators prove a more suitable choice. Phase modulators, which are standard components in optical communications, offer wavelength transparency, high radio-frequency bandwidth, and simplicity of operation, because they require only a single radio-frequency input and are optically linear²⁰.

Yet because phase modulators are typically driven with sinusoidal voltages, which are only approximately quadratic over a small temporal window, they suffer from severe temporal aberrations²¹. Such distortions are particularly harmful in implementing a temporal cloak, because a continuous-wave input necessarily extends well beyond the parabolic peaks of the sinusoid. Moreover, the original temporal cloak uses split time-lenses, which apply a discontinuous frequency chirp to the continuous-wave probe; after propagating through dispersive fibre, the spectral content of the waveform separates in time, leaving a gap with zero intensity¹⁵. This discontinuous chirp requires that the parabolic approximation remain valid all the way to the edges of the time lens—precisely where it breaks down completely for a sinusoid. Further, even if we generate a non-sinusoidal radio-frequency signal that more accurately approximates a parabola, replicating the chirp discontinuity still requires extremely high bandwidth at repetition rates suitable for telecommunications. (See the Supplementary Information for further discussion.) Under these restrictions, an alternative to the split time-lens is required for temporal cloaking in the gigahertz regime.

Interestingly, the desired transformation of continuous-wave light into clean, high-extinction pulses is closely related to the generation of optical frequency combs through electro-optic modulation. In this application, the goal is to convert a continuous-wave input into a broadband frequency comb with a smooth spectrum. One such method for flat comb generation exploits a temporal version of the Talbot effect. Observed in spatial optics as early as 1836, the Talbot phenomenon yields perfect regeneration of the optical field at discrete distances away from a periodic grating²². Through space-time duality, a temporal analogue arises²³. Specifically, for an electric-field envelope periodic at the radio frequency ω_{rep} , where ‘rep’ indicates repetition frequency, modulated by an optical carrier at frequency ω_0 and traversing a medium described by propagation constant $\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2$, the waveform exactly reproduces itself at multiples of the Talbot distance $L_T = 4\pi/|\beta_2|\omega_{\text{rep}}^2$.

As a special case, when a continuous-wave input is sinusoidally phase-modulated at an amplitude of $\pi/4$, then propagated through the fractional Talbot distance $L_T/4$, the output waveform consists of high-extinction, 50% duty-cycle pulses^{24–27}. These pulses can be imaged effectively by a second application of sinusoidal phase, because at this point the optical energy lies primarily within a window where the quadratic phase of an ideal time lens is well approximated. Subsequent dispersion chosen to satisfy the temporal imaging condition compresses these pulses even further, thereby creating large temporal gaps, the hallmark of a cloak. And this compression is achieved with phase-only elements—which are reversible apart from linear insertion loss—so it can be undone with inverse dispersion and modulation, thus completing the temporal cloak. The spatial equivalent of this cloaking circuit is highlighted in Fig. 1, revealing the large cloaking window possible with the Talbot effect. We emphasize that no discontinuity in the chirp rate is required.

Figure 2a presents the full experimental arrangement. The first phase modulator applies a small phase modulation to the input, and a chirped fibre Bragg grating provides the required fractional Talbot dispersion. The second phase modulator widens the signal bandwidth, and optical fibre compresses the waveform in time. The spectro-temporal characteristics of the optical probe at the event plane are summarized in Fig. 2b and c; the broadband frequency comb is compressed smoothly to an autocorrelation full-width at half-maximum (FWHM) of 11.7 ps, corresponding to about 15% of the 78.7-ps repetition period. The following fibre, phase modulators, and chirped fibre Bragg grating simply undo the effects of their counterparts, leaving a continuous-wave output. The extra dispersive link after the final phase modulator ensures that the cloak can itself be hidden; that is, when the phase modulators are switched off, the applied event appears at the output unaltered, as if the cloak were absent completely. This requires that a net dispersion of around 0 ps nm^{-1} be experienced by the uncloaked event, which necessitates the additional dispersive link. On the other hand, when the cloak is operational, the output of the last phase modulator is essentially continuous-wave, so the extra dispersion has no impact. As another modification, we employ a 12.3-Gb s^{-1} photoreceiver for detection of the temporal output, presenting bandwidth filtering as a cloak enhancer. For relatively narrowband events, a spectral filter with properly chosen bandwidth can be used to remove

¹School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA.

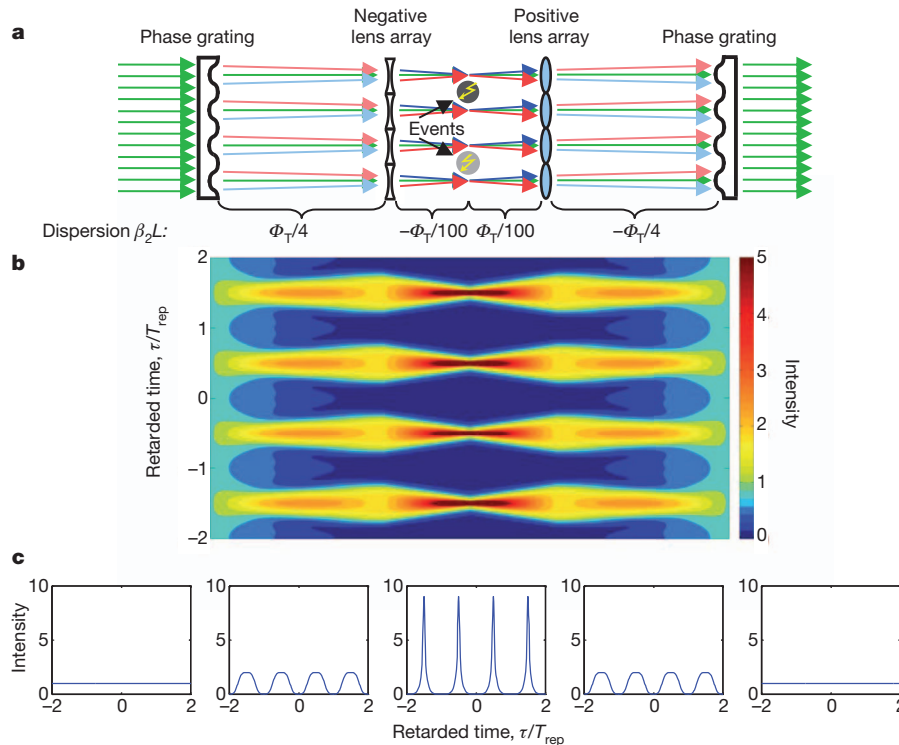


Figure 1 | Spatial analogue of temporal cloaking circuit. **a**, Temporal ray diagram highlighting the spatial equivalent of the experimental set-up. $\Phi_T = \beta_2 L_T$ represents the Talbot dispersion for dispersion-compensating fibre with dispersion constant β_2 . Owing to the diffractive nature of the Talbot effect, temporal ray optics is not strictly applicable, but we nonetheless include this ray

diagram for visualization. **b**, Corresponding simulated intensity distribution. Wide cloaking windows of zero intensity appear at the temporal focus, nearing the duration of the repetition period, T_{rep} . **c**, Temporal intensity slices at specific locations in the circuit (panels from left to right): first grating, negative lens array, event plane, positive lens array and final grating.

residual high-frequency sidebands—resulting from cloak imperfections—while still passing the event itself intact. This principle, discussed in detail in the Supplementary Information, proves extremely useful in cloak operation. Under these conditions, the reconstructed waveform is as summarized in Fig. 2d and e. The cloak is able to reproduce the continuous-wave input: the final spectrum consists of one line, matching the input spectrum, and the temporal waveform is nearly flat, albeit with some parasitic modulation due to cloak imperfections.

Applying a sinusoid to the electro-optic intensity modulator in the event plane, we obtain the results of Fig. 3. We find that the cloak completely hides the presence of the perturbation, removing the spectral sidebands and turning the high-contrast temporal modulation into a nearly flat line. This periodic event enables measurement of a defining metric of cloak performance: the cloaking window. To quantify this aspect, we look at the photodetected signal's relative root-mean-square fluctuation as the perturbation is shifted in time from the optimum cloaking point. For definiteness, the cloaking window is taken as the temporal offset at which this fractional modulation has increased to one-half the value in the uncloaked case; Fig. 3d furnishes the results of this measurement. A cloaking window of 46% is found, which represents a conservative estimate: the sinusoidal modulation is of significant duration itself, so the actual cloaked region is wider than that indicated simply by the temporal offset. Unlike the arrangement in ref. 15, which can be viewed as a cloak of temporally isolated events, the periodicity in ours cannot be ignored; in fact, it is precisely this periodicity which permits use of the Talbot effect. In this sense, it is profitable also to compare our temporal cloak to metamaterial cloaking arrays²⁸. In a recent experiment²⁹ using tapered gold-coated waveguides, about 20% of the total two-dimensional surface area was cloaked—a number similar to what is obtained here. Thus our cloak meets at the intersection of two recent metamaterial concepts: the temporal cloak and the cloaking array.

In addition to hiding a deterministic periodic signal, our cloak is also able to mask pseudorandom data. We use an inverted (dark)

return-to-zero modulation format, which ensures that the optical transmission function returns to a maximum during each cycle, as required to provide temporal regions through which the compressed probe pulses can pass. Cloak performance for dark return-to-zero data is summarized in Fig. 4, for both pseudorandom and specific bit sequences. When the phase modulators are off, high-contrast voltage transitions are evident; when the cloak is turned on, these transitions reduce to a single flat line, and the data are effectively cloaked. This temporal cloak consequently succeeds in hiding communications at will, by simply turning four phase modulators on and off.

Moreover, future cloaks based on our arrangement have the potential for significant improvements, both in terms of operational bandwidth and the duration of the cloaked region. Two distinct bandwidths deserve consideration: that of the input probe, and that of the event to be cloaked. Concerning the probe, the cloak is fundamentally narrow-band. The dispersion and phase modulation are selected precisely under the assumption of a continuous-wave optical input, and a broad-band optical input is not guaranteed to develop into sharp pulses at the event plane. On the other hand, the bandwidth of the event could in principle be made much wider. Because of the filtering effect of the detection scheme (see Supplementary Information), the current cloak is admittedly limited to event bandwidths approximately twice the modulation frequency. However, this filtering is required only because of imperfections in the cloak itself, particularly the difficulty in exactly matching two phase modulators. With improved uniformity in the phase modulators, such filtering could be removed, permitting disturbances with much broader frequency content. In fact, the event could possess a bandwidth wider than that created by the phase modulation for probe compression, provided it lies within the passband of the optical components used. The disturbance must be temporally restricted to allow unity transmission over some fraction of the period, but no such restriction is imposed on its modulation bandwidth. Indeed, impulsive events with extremely large bandwidths are actually easier

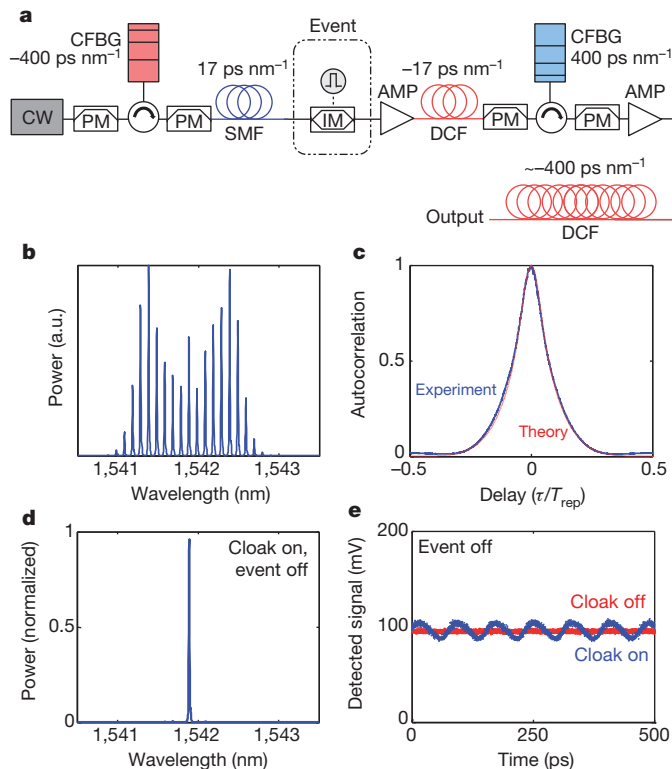


Figure 2 | Experimental set-up. **a**, Schematic of the complete cloaking circuit. CW, continuous-wave input laser; PM, phase modulator; CFBG, chirped fibre Bragg grating; SMF, single-mode fibre; DCF, dispersion-compensating fibre; IM, intensity modulator; AMP, erbium-doped fibre amplifier. **b**, Comb spectrum at the event plane, consisting of 16 spectral lines in the 10-dB bandwidth. **c**, Corresponding intensity autocorrelation, shown over one full temporal period. The FWHM is 11.7 ps. **d**, Spectrum at the output of the cloaking circuit, when no event is applied. **e**, Corresponding temporal output measured on a photodetector, compared to the case when all phase modulators are off.

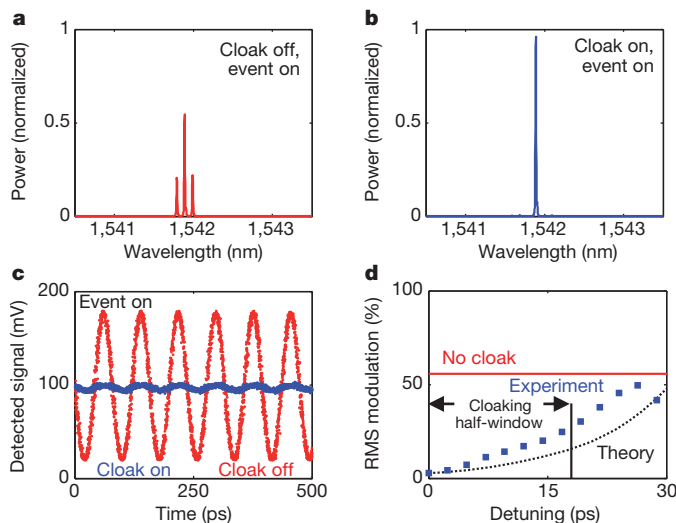


Figure 3 | Cloaking of sinusoidal modulation. **a**, Output spectrum when the phase modulators are off and a sinusoid is applied to the intensity modulator. **b**, Spectrum when the cloak is on, demonstrating removal of the sidebands in **a**. (Spectra are normalized as in Fig. 2d.) **c**, Corresponding temporal output. When the cloak is turned on, the previously high-contrast modulation is reduced to a flat line, hiding this event from an observer. **d**, Measurement of the temporal cloaking window. The fractional modulation reaches one-half that in the uncloaked case at a detuning of 18 ps, for a total double-ended cloaking window of 36 ps, or 46% of the temporal period.

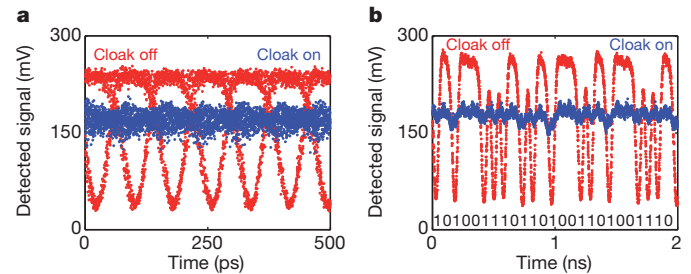


Figure 4 | Cloaking of data. **a**, Temporal output when length $2^{31} - 1$ pseudorandom data are applied to the intensity modulator, measured on a sampling oscilloscope. The clear transitions between high and low data levels present when the cloak is off are completely removed when the phase modulators are on. **b**, Output for a particular sequence of ones and zeros. Although the binary data specified on the bottom of the plot are clearly detected when the cloak is off, the voltage swings indicative of bit transmission are suppressed to a nearly flat line when the cloak is on.

to cloak than the data examined here, because they fit easily within the cloaking window. Additionally, from an operational perspective, the data rate of 12.7 GHz could be easily increased to 40 GHz and higher with state-of-the-art LiNbO₃ modulators. Because the required values of optical dispersion depend on this choice of frequency, one can simply choose different dispersive links to convert our cloak at 12.7 GHz to any other convenient repetition frequency.

The 46% cloaking window is currently limited by phase modulator performance. The duration of the compressed probe beam (Fig. 2c) is inversely proportional to the optical bandwidth after the second phase modulator in our experiment (spectrum in Fig. 2b), which is in turn directly proportional to the phase modulator's modulation index²⁷. The modulation index itself is limited by the phase modulator's maximum allowable radio-frequency input power, which prevents much shorter probe pulses and wider cloaking windows with the current arrangement. Yet by replacing the second and third phase modulators each with separate series of cascaded modulators, the net effective index can be increased even without improved technology. In fact, this principle has already been applied in pulse compression studies³⁰, yielding pulses more than eight times shorter than what we obtain here. This implies that a fractional cloaking window of over 90% could be possible in our set-up, using three cascaded phase modulators instead of one. Closely approaching the limit of 100% would at present require too many phase modulators to be practical, but it nonetheless remains a possibility for the future; nothing inherently prevents it.

From a more fundamental perspective, our experiments highlight the efficacy of the Talbot effect for general cloaking. In a sense, it produces its own cloak, even in a single dispersive medium: through self-imaging, a phase-modulated signal that is compressed to pulses at $L_T/4$ will naturally return to continuous-wave light upon further propagation. Therefore, although we used additional time-lensing to substantially expand the cloaking window, this is not intrinsically necessary. One could envision removing these time lenses entirely and simply letting the Talbot effect run its course. A spatial equivalent could then be implemented with phase gratings in a simple uniform medium. This example stresses yet again the unique insights afforded by space-time duality for enriching our understanding of seemingly disparate phenomena.

METHODS SUMMARY

All phase modulators are driven by a single low-noise sine-wave generator (Agilent E8257D) operating at a frequency of 12.71 GHz. The gain of all radio-frequency amplifiers is regulated to ensure matched modulation, either via a control voltage or with tunable radio-frequency attenuators. The first and fourth phase modulators are driven at a modulation index of $\pi/4$, and the second and third at an index of around 2π , limited by the maximum allowable radio-frequency input power. The continuous-wave laser (Koheras AdjustiK) is operated at about 1,541.9 nm, and radio-frequency phase shifters are used to align the applied phase of each phase modulator. The optical fibre consists of around a kilometre of

standard Corning SMF-28e and a dispersion-compensating fibre module from Optical Fibre Solutions, with net dispersions of around 17 ps nm^{-1} and -17 ps nm^{-1} , respectively; the chirped fibre Bragg gratings were designed by Proximion Fibre Systems, and feature approximate dispersions of $\pm 400 \text{ ps nm}^{-1}$. Erbium-doped fibre amplifiers are required to compensate for the insertion loss of the optical components.

For sinusoidal modulation, the intensity modulator is driven with the clock signal directly and is biased at the half-power point; for data, a bit-error-rate tester (Agilent N4901B) generates a non-return-to-zero sequence which is then converted to the desired return-to-zero drive signal by a digital logic circuit (Hittite HMC706LC3C) and applied to the intensity modulator, this time at zero bias. A fraction of the final output is split off and recorded on an optical spectrum analyser with a resolution of 0.01 nm . The remainder is detected with a 12.3-Gb s^{-1} photoreceiver (Agere 2560A-C02), and the electrical output is measured on a fast sampling oscilloscope (Tektronix DSA8200). The voltage levels in the cloaked cases do not align with the uncloaked peaks because of amplifier saturation, which forces the average output power to remain constant; a continuous-wave waveform must drop slightly compared to the modulated case to conserve the integrated power.

Currently, thermo-optic effects in the optical fibre links limit long-term operation. Small changes in the refractive index resulting from thermal drift cause the timing between complementary modulators to lose synchronization; under the conditions in our laboratory, this means that the phase shifters require slight readjustment approximately every 15 min. However, replacing all optical fibre with equivalent chirped fibre Bragg gratings would remove this instability by lowering the total length of silica from over a kilometre to just a few metres, thereby making timing drift negligible relative to the radio-frequency period.

Received 18 February; accepted 12 April 2013.

Published online 5 June 2013.

- Veselago, V. G. The electrodynamics of substances with simultaneously negative values of ϵ and μ . *Sov. Phys. Usp.* **10**, 509 (1968).
- Pendry, J. B., Holden, A. J., Robbins, D. J. & Stewart, W. J. Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Trans. Microw. Theory Tech.* **47**, 2075–2084 (1999).
- Smith, D. R., Padilla, W. J., Vier, D. C., Nemat-Nasser, S. C. & Schultz, S. Composite medium with simultaneously negative permeability and permittivity. *Phys. Rev. Lett.* **84**, 4184–4187 (2000).
- Alù, A. & Engheta, N. Achieving transparency with plasmonic and metamaterial coatings. *Phys. Rev. E* **72**, 016623 (2005).
- Leonhardt, U. Optical conformal mapping. *Science* **312**, 1777–1780 (2006).
- Pendry, J. B., Schurig, D. & Smith, D. R. Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
- Schurig, D. *et al.* Metamaterial electromagnetic cloak at microwave frequencies. *Science* **314**, 977–980 (2006).
- Li, J. & Pendry, J. B. Hiding under the carpet: a new strategy for cloaking. *Phys. Rev. Lett.* **101**, 203901 (2008).
- Shalaev, V. M. Transforming light. *Science* **322**, 384–386 (2008).
- Liu, R. *et al.* Broadband ground-plane cloak. *Science* **323**, 366–369 (2009).
- Valentine, J., Li, J., Zentgraf, T., Bartal, G. & Zhang, X. An optical cloak made of dielectrics. *Nature Mater.* **8**, 568–571 (2009).
- Smolyaninov, I. I., Smolyaninova, V. N., Kildishev, A. V. & Shalaev, V. M. Anisotropic metamaterials emulated by tapered waveguides: application to optical cloaking. *Phys. Rev. Lett.* **102**, 213901 (2009).
- Chen, H., Chan, C. T. & Sheng, P. Transformation optics and metamaterials. *Nature Mater.* **9**, 387–396 (2010).
- McCall, M. W., Favaro, A., Kinsler, P. & Boardman, A. A spacetime cloak, or a history editor. *J. Opt.* **13**, 024003 (2011).
- Fridman, M., Farsi, A., Okawachi, Y. & Gaeta, A. L. Demonstration of temporal cloaking. *Nature* **481**, 62–65 (2012).
- Kolner, B. H. & Nazarathy, M. Temporal imaging with a time lens. *Opt. Lett.* **14**, 630–632 (1989).
- Kolner, B. H. Space-time duality and the theory of temporal imaging. *IEEE J. Quantum Electron.* **30**, 1951–1963 (1994).
- Bennett, C. V. & Kolner, B. H. Principles of parametric temporal imaging. I. System configurations. *IEEE J. Quantum Electron.* **36**, 430–437 (2000).
- Bennett, C. V. & Kolner, B. H. Principles of parametric temporal imaging. II. System performance. *IEEE J. Quantum Electron.* **36**, 649–655 (2000).
- Torres-Company, V., Lancis, J. & Andrés, P. in *Progress in Optics* (ed. Wolf, E.) Vol. 56, Ch. 1 (Elsevier, 2011).
- Bennett, C. V. & Kolner, B. H. Aberrations in temporal imaging. *IEEE J. Quantum Electron.* **37**, 20–32 (2001).
- Patorski, K. in *Progress in Optics* (ed. Wolf, E.) Vol. 27, Ch. 1 (Elsevier, 1989).
- Jannson, T. & Jannson, J. Temporal self-imaging effect in single-mode fibers. *J. Opt. Soc. Am.* **71**, 1373–1376 (1981).
- Guigay, J. P. On Fresnel diffraction by one-dimensional periodic objects, with application to structure determination of phase objects. *Opt. Acta* **18**, 677–682 (1971).
- Komukai, T., Yamamoto, T. & Kawanishi, S. Optical pulse generator using phase modulator and linearly chirped fiber Bragg gratings. *IEEE Photon. Technol. Lett.* **17**, 1746–1748 (2005).
- Torres-Company, V., Lancis, J. & Andrés, P. Unified approach to describe optical pulse generation by propagation of periodically phase-modulated CW laser light. *Opt. Express* **14**, 3171–3180 (2006).
- Torres-Company, V., Lancis, J. & Andrés, P. Lossless equalization of frequency combs. *Opt. Lett.* **33**, 1822–1824 (2008).
- Farhat, M. *et al.* Understanding the functionality of an array of invisibility cloaks. *Phys. Rev. B* **84**, 235105 (2011).
- Smolyaninova, V. N., Smolyaninov, I. I. & Ermer, H. K. Experimental demonstration of a broadband array of invisibility cloaks in the visible frequency range. *New J. Phys.* **14**, 053029 (2012).
- Metcalfe, A. J., Torres-Company, V., Leaird, D. E. & Weiner, A. M. High-power broadly tunable electro-optic frequency comb generator. *IEEE J. Sel. Top. Quant. Electron.* (in the press).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank V. Shalaev and V. Torres-Company for comments and discussions. This project was supported in part by the National Science Foundation (grant number ECCS-1102110) and the Naval Postgraduate School (grant number N00244-09-1-0068) under the National Security Science and Engineering Faculty Fellowship programme. (Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsors.) J.M.L. acknowledges financial support from the Department of Defense through a National Defense Science and Engineering Graduate Fellowship.

Author Contributions J.M.L. conducted the experiments, with assistance from D.E.L. A.M.W. initiated and supervised the project. All authors discussed the results and the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.W. (amw@purdue.edu).

Heat dissipation in atomic-scale junctions

Woochul Lee^{1*}, Kyeongtae Kim^{1*}, Wonho Jeong¹, Linda Angela Zotti², Fabian Pauly³, Juan Carlos Cuevas² & Pramod Reddy^{1,4}

Atomic and single-molecule junctions represent the ultimate limit to the miniaturization of electrical circuits¹. They are also ideal platforms for testing quantum transport theories that are required to describe charge and energy transfer in novel functional nanometre-scale devices. Recent work has successfully probed electric and thermoelectric phenomena^{2–8} in atomic-scale junctions. However, heat dissipation and transport in atomic-scale devices remain poorly characterized owing to experimental challenges. Here we use custom-fabricated scanning probes with integrated nanoscale thermocouples to investigate heat dissipation in the electrodes of single-molecule ('molecular') junctions. We find that if the junctions have transmission characteristics that are strongly energy dependent, this heat dissipation is asymmetric—that is, unequal between the electrodes—and also dependent on both the bias polarity and the identity of the majority charge carriers (electrons versus holes). In contrast, junctions consisting of only a few gold atoms ('atomic junctions') whose transmission characteristics show weak energy dependence do not exhibit appreciable asymmetry. Our results unambiguously relate the electronic transmission characteristics of atomic-scale junctions to their heat dissipation properties, establishing a framework for understanding heat dissipation in a range of mesoscopic systems where transport is elastic—that is, without exchange of energy in the contact region. We anticipate that the techniques established here will enable the study of Peltier effects at the atomic scale, a field that has been barely explored experimentally despite interesting theoretical predictions^{9–11}. Furthermore, the experimental advances described here are also expected to enable the study of heat transport in atomic and molecular junctions—an important and challenging scientific and technological goal that has remained elusive^{12,13}.

Charge transport is always accompanied by heat dissipation (Joule heating). This process is well understood at the macroscale, where the power dissipation (heat dissipated per unit time) is volumetric and is given by $j^2\rho$, where j is the magnitude of the current density and ρ is the electrical

resistivity. Heating in atomic-scale junctions is expected to be fundamentally different, as charge transport through such junctions is largely elastic^{14,15}. Recent experiments have probed the local non-equilibrium electronic and phononic temperatures in molecular junctions^{16–18} to obtain insights into the effect of electron–electron and electron–phonon interactions on heat dissipation at the atomic scale. However, experimental challenges in quantitatively measuring atomic-scale heat dissipation have impeded the elucidation of a fundamental question: what is the relationship between the electronic transmission characteristics of atomic and molecular junctions (AMJs) and their heat dissipation properties?

In this work, we overcome this challenging experimental hurdle by using custom-fabricated nanoscale–thermocouple integrated scanning tunnelling probes (NTISTPs; Fig. 1a and b). The NTISTPs feature an outer gold (Au) electrode that is electrically isolated but thermally well connected to the integrated gold–chromium thermocouple via a thin (70 nm) silicon nitride film (see Supplementary Information for fabrication details). To probe heat dissipation, we first created a series of AMJs (Fig. 1c) between the outer Au electrode of the NTISTP and a flat Au substrate. Application of a voltage bias across such AMJs results in a temperature rise of the integrated thermocouple due to heat dissipation in the NTISTP's apex on a length scale comparable to the inelastic mean free path of electrons in Au (ref. 19). The power dissipation in the probe (Q_P) and the temperature rise of the thermocouple (ΔT_{TC}), located ~ 300 nm away from the apex, are directly related by $Q_P = \Delta T_{TC}/R_P$ (see Methods), where R_P is the thermal resistance of the NTISTP (see Fig. 1b). Further, ΔT_{TC} is related to the thermoelectric voltage output of the thermocouple (ΔV_{TC}) by $\Delta V_{TC} = -S_{TC}\Delta T_{TC}$, where S_{TC} is the effective Seebeck coefficient of the thermocouple. We note that R_P and S_{TC} were experimentally determined to be $72,800 \pm 500$ K W^{−1} and 16.3 ± 0.2 μ V K^{−1}, respectively (Supplementary Information).

We began our experimental studies, at room temperature, by trapping single molecules of 1,4-benzenediisonitrile (BDNC; Fig. 1c) between the Au electrodes of the NTISTP and the substrate using a break junction technique^{5,20}. We first obtained electrical conductance

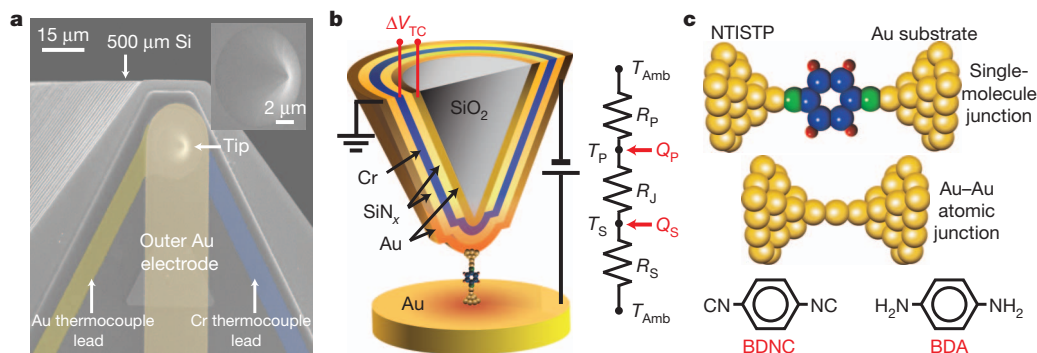


Figure 1 | Nanoscale thermocouple probes and atomic and molecular junctions studied in this work. **a**, Scanning electron microscope (SEM) image of a NTISTP. The electrodes are false-coloured. Inset, magnified image of the tip. **b**, Diagram of a junction created between the NTISTP (cross-sectional

view) and a Au substrate (bottom) along with a thermal resistance network (right) that represents the dominant resistances to heat flow. **c**, Diagrams of molecular and atomic junctions (top) along with the structures of the molecules studied (bottom). (All diagrams are not drawn to scale or proportion.)

¹Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA. ²Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, Madrid 28049, Spain. ³Department of Physics, University of Konstanz, D-78457 Konstanz, Germany. ⁴Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA.

*These authors contributed equally to this work.

versus displacement traces by monitoring the electrical current under an applied bias while the NTISTP–substrate separation was systematically varied. Figure 2a shows representative conductance traces along with a histogram obtained from 500 such curves. The histogram features a peak at $\sim 0.002G_0$ ($G_0 = 2e^2/h \approx (12.9 \text{ k}\Omega)^{-1}$), which represents the most probable low-bias conductance of Au–BDNC–Au junctions ($G_{\text{Au-BDNC-Au}}$) and is in good agreement with past work²¹.

To probe heat dissipation, we created stable Au–BDNC–Au junctions with a conductance that is within 10% of the most probable low-bias conductance²⁰. We studied heat dissipation in 100 distinct Au–BDNC–Au junctions, at each bias, to obtain the time-averaged temperature rise ($\Delta T_{\text{TC,Avg}}$) and the time-averaged power dissipation in the NTISTP ($Q_{\text{P,Avg}}$) for both positive and negative biases. Here, a positive (negative) bias corresponds to a scenario where the probe is grounded, while the substrate is at a higher (lower) potential. We note that a modulated voltage bias was applied to the junctions to obtain $\Delta T_{\text{TC,Avg}}$ —with high resolution—for both positive and negative biases (see Methods and Supplementary Information). This modulation scheme enables rejection of broadband noise and plays a critical role in performing high-resolution thermometry.

The circles (triangles) in Fig. 2b represent the measured $\Delta T_{\text{TC,Avg}}$ as well as the estimated $Q_{\text{P,Avg}}$ for positive (negative) biases as a function of the time-averaged total power dissipation in the junctions ($Q_{\text{Total,Avg}}$) at each bias voltage. Here, $Q_{\text{Total,Avg}}$ represents all the power dissipated in the junction, at a given bias voltage, and can be readily obtained from the measured current (I) and the known voltage bias (V) applied to the junction (see Methods). We note that the current–voltage (I – V) characteristics of Au–BDNC–Au junctions are nonlinear (Fig. 2c), therefore, in general $Q_{\text{Total,Avg}} \neq G_{\text{Au-BDNC-Au}} V^2$. The dotted line in Fig. 2b corresponds to the expected temperature rise of the probe if the heating was symmetric, that is, if half of the total power was dissipated in the probe ($\Delta T_{\text{Symmetric}} = Q_{\text{Total,Avg}}/2R_{\text{P}}$). It can be clearly seen that for a given $Q_{\text{Total,Avg}}$ the power dissipation in

the probe is larger under a negative bias than a positive bias. We also conclude that the time-averaged power dissipation in the substrate, $Q_{\text{S,Avg}}$, is smaller under a negative bias than under a positive bias, because $Q_{\text{P,Avg}} + Q_{\text{S,Avg}} = Q_{\text{Total,Avg}}$. To clarify the voltage biases used in the experiments, we present (in the inset of Fig. 2b) $\Delta T_{\text{TC,Avg}}$ as a function of the magnitude of the applied voltage bias. These results unambiguously demonstrate that heat dissipation in the electrodes of Au–BDNC–Au junctions is bias polarity dependent and unequal.

This observation raises an important question: why is the heat dissipation in the electrodes unequal in spite of the symmetric geometry of the molecular junctions? To address this question, we resort to the Landauer theory of quantum transport, which has successfully described charge transport in numerous nanostructures¹⁹. Within this theory, the power dissipated in the probe and the substrate, $Q_{\text{P}}(V)$ and $Q_{\text{S}}(V)$, respectively, is given by²²:

$$Q_{\text{P}}(V) = \frac{2}{h} \int_{-\infty}^{\infty} (\mu_{\text{P}} - E) \tau(E, V) [f_{\text{P}} - f_{\text{S}}] dE$$

$$Q_{\text{S}}(V) = \frac{2}{h} \int_{-\infty}^{\infty} (E - \mu_{\text{S}}) \tau(E, V) [f_{\text{P}} - f_{\text{S}}] dE \quad (1)$$

Here μ_{P} and μ_{S} are the chemical potentials of the probe and substrate electrodes, respectively, $f_{\text{P/S}}$ represent the Fermi–Dirac distribution of the probe/substrate electrodes, and $\tau(E, V)$ is the energy (E) and voltage bias (V) dependent transmission function. Equation (1) suggests that the power dissipation in the two electrodes is, in general, unequal, that is, $Q_{\text{P}}(V) \neq Q_{\text{S}}(V)$, and bias polarity dependent, that is, $Q_{\text{P/S}}(V) \neq Q_{\text{P/S}}(-V)$. Specifically, it is straightforward to show that:

$$Q_{\text{P}}(V) - Q_{\text{P}}(-V) \approx 2G_{\text{TSV}} + O(V^3)$$

$$Q_{\text{P}}(V) - Q_{\text{S}}(V) \approx 2G_{\text{TSV}} + O(V^3) \quad (2)$$

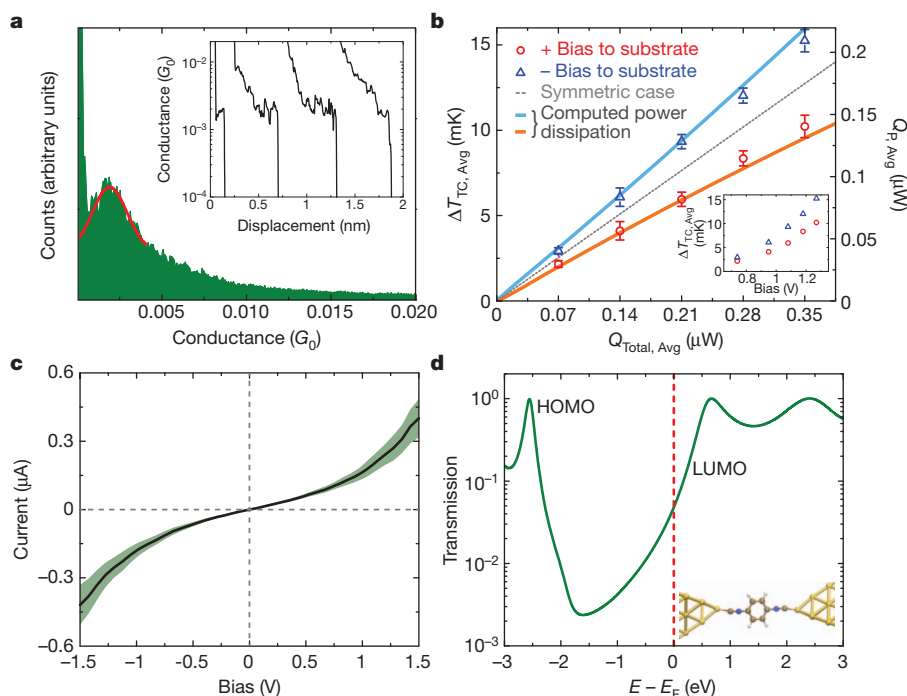


Figure 2 | Relationship between heat dissipation asymmetries and electronic transmission characteristics in Au–BDNC–Au junctions.

a, Horizontally offset conductance traces (inset) of BDNC junctions, along with a histogram obtained from 500 traces (main panel). The red line represents a Gaussian fit to the histogram. **b**, Main panel, measured time-averaged temperature rise of the thermocouple ($\Delta T_{\text{TC,Avg}}$) and the time-averaged power dissipation in the probe ($Q_{\text{P,Avg}}$) as a function of the time-averaged total power dissipation in the junction ($Q_{\text{Total,Avg}}$) for positive and negative biases. Error bars represent the estimated uncertainty in $\Delta T_{\text{TC,Avg}}$ (see Supplementary Information for details of uncertainty estimation). The computationally predicted relationship between Q_{P} and Q_{Total} is shown by solid lines, which illustrates that $Q_{\text{P}} = fQ_{\text{Total}}$, where f is dependent on both Q_{Total} and the polarity of the applied bias, and is in general not equal to 0.5. The dotted line corresponds to the expected temperature rise of the probe if the heating was symmetric (that is, $f = 0.5$). Inset, measured $\Delta T_{\text{TC,Avg}}$ as a function of the magnitude of the applied voltage bias. Uncertainties are not shown in the inset, for visual clarity. **c**, I – V characteristics of Au–BDNC–Au junctions obtained by averaging 100 individual I – V curves (solid curve). The shaded region represents the standard deviation of the I – V curves. **d**, Computed zero-bias transmission function corresponding to the Au–BDNC–Au junction shown in the inset. HOMO, highest occupied molecular orbital; LUMO, lowest unoccupied molecular orbital.

Here, G is the low bias electrical conductance of the junction, T is the absolute temperature, and S is the Seebeck coefficient of the junction, whose sign is related to the first energy derivative of the zero-bias transmission $\tau'(E = E_F, V = 0)$ at the Fermi energy (E_F), resulting in a positive Seebeck coefficient for a negative first derivative and vice versa²³. To test if the observed heating asymmetry can be understood within this framework, we computed $\tau(E, V = 0)$ for Au–BDNC–Au junctions using a transport method²⁴ based on density functional theory (DFT; Methods). The computed transmission function (Fig. 2d) exhibits a positive slope at the Fermi energy, in agreement with past work²⁵, indicating a negative Seebeck coefficient, which by virtue of equation (2) leads to higher power dissipation in the NTISTP when negative voltages are applied to the substrate. Further, the solid lines in Fig. 2b represent the relationship between Q_P and Q_{Total} ($Q_P + Q_S = Q_{\text{Total}}$) as computed from equation (1) under the assumption that $\tau(E, V)$ is well approximated by $\tau(E, V = 0)$. Notice that although our DFT approach overestimates the linear conductance, it describes correctly the relationship between Q_P and Q_{Total} . The reasons for this agreement are discussed further in the Supplementary Information, where we show in particular that this relation is relatively insensitive to the details of the junction geometry. The good agreement of the computed and measured relation between power dissipations provides strong support to the applicability of the Landauer theory of heat dissipation at the atomic scale.

To prove conclusively the relationship between electronic structure and heat dissipation, we performed additional studies on 1,4-benzenediamine (BDA; Fig. 1c) junctions, which are expected to exhibit hole-dominated electrical transport, as suggested by our calculations (Fig. 3d) and past experiments²⁶. Following a procedure similar to that described above, we first determined that the most probable low-bias conductance of Au–BDA–Au junctions was $\sim 0.005G_0$ (Fig. 3a), a value consistent with past work²⁷. Measurements of heat dissipation in BDA junctions (Fig. 3b) show a remarkably different asymmetry. In particular, the BDA junctions show larger power dissipation in the probe for a positive bias than for a negative one—in strong contrast to that observed in BDNC junctions. To understand this important difference, we computed the transmission function of the Au–BDA–Au junction displayed in Fig. 3d, which shows that $\tau'(E = E_F, V = 0)$ is negative,

resulting in a positive Seebeck coefficient. This, in turn, leads to larger power dissipation in the NTISTP at positive biases. Further, the computed relationship between Q_P and Q_{Total} is in good agreement with our experimental observations (solid lines in Fig. 3b).

Finally, to prove that no appreciable asymmetries are obtained if the transmission is weakly dependent on energy, we studied heat dissipation in Au–Au atomic junctions. We began our analysis by studying the conductance of Au–Au atomic junctions, which were found to have a most probable conductance of $\sim G_0$, in accordance with past studies^{5,28} (Supplementary Information). Subsequently, we created 100 Au–Au atomic junctions with a low-bias conductance of $G_0 \pm 0.1G_0$ and probed heating in them. The measured $\Delta T_{\text{TC, Avg}}$ (Fig. 4a) is seen to be proportional to $Q_{\text{Total, Avg}}$ and is identical for both positive and negative biases (within experimental uncertainty, ~ 0.1 mK), clearly demonstrating that there is no detectable asymmetry in the power dissipation. Further, additional experiments performed at larger values of $Q_{\text{Total, Avg}}$ also show no detectable asymmetry (Fig. 4a inset).

Symmetric heat dissipation is indeed expected in Au–Au atomic junctions because of the weak energy dependence of their transmission function²⁹, which is reflected in the fact that their average thermopower vanishes⁶. In Fig. 4b we present the computed zero-bias transmission, corresponding to the Au–Au atomic junction shown in the left inset. The transmission is practically energy independent over 1 eV around the Fermi energy. This weak energy dependence results in symmetric power dissipation (from equations (1) and (2)) as well as linear I – V characteristics, as evidenced by the experimentally obtained I – V curves shown in the right inset of Fig. 4b.

The good agreement between the measured and computed asymmetries in the heat-dissipation characteristics of AMJs unambiguously confirms that heat dissipation is indeed intimately related to the transmission characteristics of the junctions, as predicted by the Landauer theory. We note that our results contradict recent claims³⁰ of asymmetric heat dissipation in Au atomic junctions that are not in agreement with theoretical predictions. The insights obtained here regarding heat dissipation should hold for any mesoscopic system where charge transport is predominantly elastic. Such systems include semiconductor nanowires, two-dimensional electron gases, semiconductor heterostructures, carbon nanotubes and graphene.

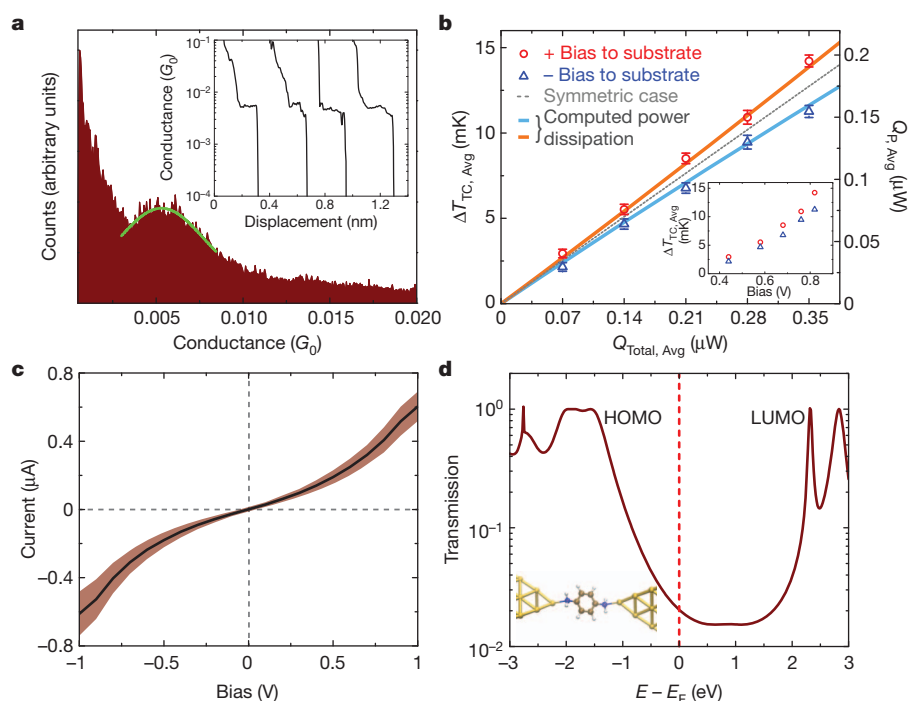


Figure 3 | Heat dissipation asymmetry for Au–BDA–Au junctions. a–d, As Fig. 2 but for Au–BDA–Au junctions. In contrast to Au–BDNC–Au junctions, the heat dissipated in the probe is found to be larger for positive biases than for negative ones.

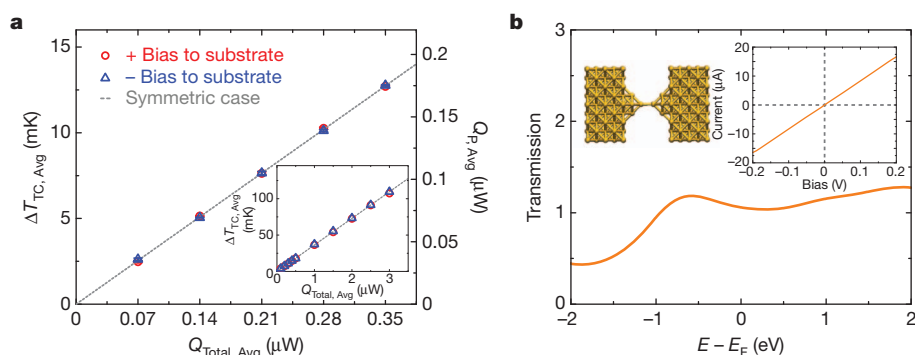


Figure 4 | No detectable heating asymmetry in Au–Au atomic junctions. **a**, The measured $\Delta T_{TC,Avg}$ and $Q_{P,Avg}$ in Au–Au atomic junctions for positive and negative biases as a function of $Q_{Total,Avg}$ (uncertainty of $\Delta T_{TC,Avg}$ is <0.1 mK for all voltage biases). Inset, results of similar measurements for a larger range of powers (uncertainty is <0.1 mK and is imperceptible in the figure). The measured temperature rise is found to be linearly dependent on $Q_{Total,Avg}$ and is independent of the bias polarity within experimental uncertainty. Further, $Q_{P,Avg} = Q_{Total,Avg}/2$ irrespective of the bias. **b**, The computed transmission function corresponding to the Au–Au atomic junction shown in the left inset features a weak energy dependence around the Fermi energy (E_F). Right inset shows the experimentally obtained I – V characteristics of Au–Au atomic junctions created by averaging over 100 independent I – V curves.

METHODS SUMMARY

Single-molecule and atomic junctions were created by displacing the NTISTP towards a Au substrate at 5 nm s^{-1} and withdrawing from the substrate at 0.1 nm s^{-1} after contact formation (indicated by an electrical conductance greater than $5G_0$). The Au substrate was coated with the desired molecules for molecular experiments and was pristine for the atomic junction studies. To obtain the conductance traces, a voltage bias of 100 mV was applied and the current was monitored during the withdrawal process. The obtained traces were analysed by creating histograms to identify the most probable conductance of AMJs. Stable single-molecule junctions with a desired conductance were created by stopping the withdrawal when a conductance plateau with a conductance within 10% of the most probable conductance was obtained. All the experiments were performed in an ultrahigh-vacuum scanning probe microscope at ambient temperature. Further, high-resolution temperature measurements were enabled by a modulation scheme where a time-dependent voltage, $V_M(t)$, consisting of a periodic series of three level voltage pulses ($+V_M$, 0 V , $-V_M$; Supplementary Fig. 1) was applied to the AMJs while monitoring the thermoelectric voltage output of the NTISTP. The zero-bias transmission functions (Figs 2–4) were computed with the *ab initio* method described in ref. 24.

Full Methods and any associated references are available in the online version of the paper.

Received 18 December 2012; accepted 6 April 2013.

- Cuevas, J. C. & Scheer, E. *Molecular Electronics: An Introduction to Theory and Experiment* (World Scientific, 2010).
- Scheer, E. *et al.* The signature of chemical valence in the electrical conduction through a single-atom contact. *Nature* **394**, 154–157 (1998).
- Song, H. *et al.* Observation of molecular orbital gating. *Nature* **462**, 1039–1043 (2009).
- Venkataraman, L., Klare, J. E., Nuckolls, C., Hybertsen, M. S. & Steigerwald, M. L. Dependence of single-molecule junction conductance on molecular conformation. *Nature* **442**, 904–907 (2006).
- Xu, B. Q. & Tao, N. J. Measurement of single-molecule resistance by repeated formation of molecular junctions. *Science* **301**, 1221–1223 (2003).
- Ludoph, B. & van Ruitenbeek, J. M. Thermopower of atomic-size metallic contacts. *Phys. Rev. B* **59**, 12290–12293 (1999).
- Reddy, P., Jang, S. Y., Segalman, R. A. & Majumdar, A. Thermoelectricity in molecular junctions. *Science* **315**, 1568–1571 (2007).
- Widawsky, J. R., Darancet, P., Neaton, J. B. & Venkataraman, L. Simultaneous determination of conductance and thermopower of single molecule junctions. *Nano Lett.* **12**, 354–358 (2012).
- Galperin, M., Saito, K., Balatsky, A. V. & Nitzan, A. Cooling mechanisms in molecular conduction junctions. *Phys. Rev. B* **80**, 115427 (2009).
- Dubi, Y. & Di Ventra, M. Colloquium: Heat flow and thermoelectricity in atomic and molecular junctions. *Rev. Mod. Phys.* **83**, 131–155 (2011).
- Karlström, O., Linke, H., Karlström, G. & Wacker, A. Increasing thermoelectric performance using coherent transport. *Phys. Rev. B* **84**, 113415 (2011).
- Lepri, S., Livi, R. & Politi, A. Thermal conduction in classical low-dimensional lattices. *Phys. Rep.* **377**, 1–80 (2003).
- Li, N. B. *et al.* Colloquium: Phononics: Manipulating heat flow with electronic analogs and beyond. *Rev. Mod. Phys.* **84**, 1045–1066 (2012).
- Agrait, N., Tiedt, C., Rubio-Bollinger, G. & Vieira, S. Onset of energy dissipation in ballistic atomic wires. *Phys. Rev. Lett.* **88**, 216803 (2002).
- Kim, Y., Pietsch, T., Erbe, A., Belzig, W. & Scheer, E. Benzenedithiol: a broad-range single-channel molecular conductor. *Nano Lett.* **11**, 3734–3738 (2011).

- Huang, Z. F. *et al.* Local ionic and electron heating in single-molecule junctions. *Nature Nanotechnol.* **2**, 698–703 (2007).
- Ward, D. R., Corley, D. A., Tour, J. M. & Natelson, D. Vibrational and electronic heating in nanoscale junctions. *Nature Nanotechnol.* **6**, 33–38 (2011).
- Ioffe, Z. *et al.* Detection of heating in current-carrying molecular junctions by Raman scattering. *Nature Nanotechnol.* **3**, 727–732 (2008).
- Datta, S. *Electronic Transport in Mesoscopic Systems* (Cambridge Univ. Press, 1995).
- Lee, W. & Reddy, P. Creation of stable molecular junctions with a custom-designed scanning tunneling microscope. *Nanotechnology* **22**, 485703 (2011).
- Kiguchi, M., Miura, S., Hara, K., Sawamura, M. & Murakoshi, K. Conductance of a single molecule anchored by an isocyanide substituent to gold electrodes. *Appl. Phys. Lett.* **89**, 213104 (2006).
- Sivan, U. & Imry, Y. Multichannel Landauer formula for thermoelectric transport with application to thermopower near the mobility edge. *Phys. Rev. B* **33**, 551–558 (1986).
- Paulsson, M. & Datta, S. Thermoelectric effect in molecular electronics. *Phys. Rev. B* **67**, 241403 (2003).
- Pauly, F. *et al.* Cluster-based density-functional approach to quantum transport through molecular and atomic contacts. *New J. Phys.* **10**, 125019 (2008).
- Xue, Y. Q. & Ratner, M. A. End group effect on electrical transport through individual molecules: a microscopic study. *Phys. Rev. B* **69**, 085403 (2004).
- Malen, J. A. *et al.* Identifying the length dependence of orbital alignment and contact coupling in molecular heterojunctions. *Nano Lett.* **9**, 1164–1169 (2009).
- Venkataraman, L. *et al.* Single-molecule circuits with well-defined molecular conductance. *Nano Lett.* **6**, 458–462 (2006).
- Brandbyge, M. *et al.* Quantized conductance in atom-sized wires between two metals. *Phys. Rev. B* **52**, 8499–8514 (1995).
- Nielsen, S. K. *et al.* Current-voltage curves of atomic-sized transition metal contacts: an explanation of why Au is ohmic and Pt is not. *Phys. Rev. Lett.* **89**, 066804 (2002).
- Tsutsui, M., Kawai, T. & Taniguchi, M. Unsymmetrical hot electron heating in quasi-ballistic nanocontacts. *Sci. Rep.* **2**, 217 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements P.R. acknowledges support from the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under award no. DE-SC0004871 (nanofabrication of novel scanning probes), from the National Science Foundation under award no. CBET 0844902 (instrumentation for real-time control) and from the Center for Solar and Thermal Energy conversion, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under award no. DE-SC0000957 (development of a scanning probe microscope). L.A.Z. acknowledges financial support from the Spanish MICINN through grant no. FIS2010-21883. F.P. acknowledges funding through the Carl Zeiss Stiftung, the DFG SFB 767, and the Baden-Württemberg Stiftung. P.R. thanks E. Meyhofer for discussions and comments. P.R. and J.C.C. thank A. Nitzan for discussions. J.C.C. is grateful for the hospitality provided by the Institute for Advanced Studies of the Hebrew University of Jerusalem, where part of this work was carried out.

Author Contributions The idea for the experiments was conceived by P.R. and J.C.C. The experiments were performed by W.L. and K.K. The custom-fabricated probes were designed, fabricated and characterized by K.K. and W.J. *Ab initio* charge transport calculations were performed by L.A.Z. and F.P. The manuscript was written by P.R. and J.C.C. with comments and inputs from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.R. (pramodr@umich.edu) or J.C.C. (juancarlos.cuevas@uam.es).

METHODS

Creation of atomic and molecular junctions. All the AMJs were created between a NTISTP and a Au-coated substrate by displacing the NTISTP towards a Au substrate (which was coated with the desired molecules in molecular experiments and was pristine in atomic junction experiments) at 5 nm s^{-1} and withdrawing from the substrate at 0.1 nm s^{-1} after contact formation as indicated by an electrical conductance greater than $5G_0$. To create the desired monolayers, 1 mM solutions of BDNC and BDA molecules, obtained commercially from Sigma Aldrich with a purity of $\sim 99\%$, were created in toluene/ethanol. Subsequently, a Au-coated mica substrate (electron beam evaporation) was placed in one of the solutions to allow self-assembly of molecules on the Au surface. After exposing the substrates for 12 h in a glove box filled with nitrogen gas, they were rinsed in ethanol and dried in nitrogen gas. For the experiments involving Au–Au atomic junctions, the Au-coated substrates were cleaned in ultraviolet-radiation ozone to eliminate any organic contamination on the surface. The NTISTPs were also cleaned with ultraviolet-radiation ozone in all studies and loaded into the UHV scanning probe microscope instrument. The measurement of electrical current was performed using a current amplifier (Keithley 428), whereas thermoelectric voltage measurements were performed using a voltage amplifier (Stanford Research System 560). All the data were collected at a sampling frequency of 2 kHz using a data acquisition system (National Instruments 6281). The approach, withdraw, and hold sequences were accomplished by using a real-time controller (National Instruments PXI8110).

Measurement of $\Delta T_{\text{TC,Avg}}$ using a modulation scheme. High-resolution temperature measurements are enabled by a modulation scheme in which a time-dependent voltage, $V_M(t)$, consisting of a periodic series of three level voltage pulses $+V_M$, 0 V , $-V_M$ (Supplementary Fig. 1), is applied. In all the experiments performed in this work, the period (T_P) of the voltage pulses was chosen to be $\sim 0.08 \text{ s}$ ($1/12.25 \text{ Hz}$). The selected modulation frequency is found to optimize the signal-to-noise ratio and is experimentally feasible owing to the small thermal time constants ($\sim 10 \mu\text{s}$) of the micrometre-sized NTISTPs, which enable high-fidelity tracking of temperature changes. The applied $V_M(t)$ results in both a modulated current ($I_M(t)$; see Supplementary Fig. 1) and a modulated temperature change of the thermocouple ($\Delta T_{\text{M,TC}}(t)$) due to Joule heating. Using the equation at the bottom of Supplementary Fig. 1, the time-averaged temperature rise corresponding to a positive bias $\Delta T_{\text{TC,Avg}}(+V_M)$ or a negative bias $\Delta T_{\text{TC,Avg}}(-V_M)$ can be directly related to the modulated thermoelectric voltage output ($\Delta V_{\text{M,TC}}(t)$) of the thermocouple. In probing heat dissipation in AMJs we applied the modulated voltage signal with an appropriately chosen amplitude V_M for a period of $\sim 5 \text{ s}$ to each AMJ. The resulting thermoelectric voltage signal $\Delta V_{\text{M,TC}}(t)$ was simultaneously recorded. This was repeated on ~ 100 junctions to collect data for $\sim 500 \text{ s}$

for each V_M . The obtained data were concatenated and analysed to estimate $\Delta T_{\text{TC,Avg}}$ corresponding to positive and negative biases as described above. This modulation scheme enables temperature measurements with submillikelvin resolution, as described in the Supplementary Information. The time-averaged total power dissipation ($Q_{\text{Total,Avg}}$), at each bias, was obtained by using the 500-s-long data corresponding to each V_M . Specifically, the data (measured current and known applied bias) were used to first compute the total heat dissipated at positive and negative biases. Subsequently, $Q_{\text{Total,Avg}}(+V_M/-V_M)$ was obtained by dividing the estimated total heat dissipation (corresponding to a positive or a negative bias) by the total time during which a positive bias ($+V_M$) or negative bias ($-V_M$) was applied ($\sim 500/3 \text{ s}$). The amplitudes (V_M) of the three level voltage pulses used in our studies were chosen to be 30 mV, 43 mV, 52 mV, 60 mV and 67 mV for Au–Au junctions; 0.74 V, 0.95 V, 1.08 V, 1.18 V and 1.27 V for Au–BDNC–Au junctions; and 0.44 V, 0.58 V, 0.68 V, 0.76 V and 0.82 V for Au–BDA–Au junctions. Representative traces obtained in the experiments are shown in Supplementary Information section 6.3.

Estimating $Q_{\text{P,Avg}}$ from the measured $\Delta T_{\text{TC,Avg}}$. To relate the temperature rise of the thermocouple to the time-averaged power dissipation in the probe $Q_{\text{P,Avg}}$, it is necessary to quantify the thermal resistance of the NTISTP. To elaborate, consider the resistance network shown in Fig. 1b, where the thermal resistances to heat flow in the probe (R_P), junction (R_J) and substrate (R_S) are identified. R_P was experimentally determined to be $72,800 \pm 500 \text{ K W}^{-1}$ (see Supplementary Information). The thermal resistances of AMJs (R_J) are estimated to be at least 10^7 K W^{-1} for all the AMJs studied here (see Supplementary Information for more details). Thus, $R_J \gg R_P$ and therefore $\Delta T_{\text{TC,Avg}}$ depends only on the power dissipated in the tip and is unaffected by the heating in the substrate. Thus, from a knowledge of $\Delta T_{\text{TC,Avg}}$ and R_P , the time-averaged power dissipation, $Q_{\text{P,Avg}}$, can be estimated as $Q_{\text{P,Avg}} = \Delta T_{\text{TC,Avg}}/R_P$.

Computation of the transmission function. The zero-bias transmission functions shown in the manuscript were computed with the *ab initio* method described in detail elsewhere²⁴. It is based on a combination of non-equilibrium Green's function techniques and density functional theory (DFT) and was implemented in the quantum-chemistry software package Turbomole. More details can be found in the Supplementary Information.

Computing the relationship between Q_P and Q_{Total} . We computed the power dissipated in the probe $Q_P(V)$ and the total power dissipated in the junction Q_{Total} ($Q_P(V) + Q_S(V) = Q_{\text{Total}}(V)$) using equation (1) and the zero-bias transmission curves of the molecular junctions (shown in Figs 2d and 3d). Subsequently, Q_P was plotted as a function of Q_{Total} as the relationship between Q_P and Q_{Total} is robustly predicted by our calculations (see Supplementary Information for details).

Small effect of water on upper-mantle rheology based on silicon self-diffusion coefficients

Hongzhan Fei¹, Michael Wiedenbeck², Daisuke Yamazaki³ & Tomoo Katsura¹

Water has been thought to affect the dynamical processes in the Earth's interior to a great extent. In particular, experimental deformation results^{1–4} suggest that even only a few tens of parts per million of water by weight enhances the creep rates in olivine by orders of magnitude. However, those deformation studies have limitations, such as considering only a limited range of water concentrations and very high stresses, which might affect the results. Rock deformation can also be understood as an effect of silicon self-diffusion, because the creep rates of minerals at temperatures as high as those in the Earth's interior are limited by self-diffusion of the slowest species^{5,6}. Here we experimentally determine the silicon self-diffusion coefficient D_{Si} in forsterite at 8 GPa and 1,600 K to 1,800 K as a function of water content $C_{\text{H}_2\text{O}}$ from less than 1 to about 800 parts per million of water by weight, yielding the relationship, $D_{\text{Si}} \approx (C_{\text{H}_2\text{O}})^{1/3}$. This exponent is strikingly lower than that obtained by deformation experiments (1.2; ref. 7). The high nominal creep rates in the deformation studies under wet conditions may be caused by excess grain boundary water. We conclude that the effect of water on upper-mantle rheology is very small. Hence, the smooth motion of the Earth's tectonic plates cannot be caused by mineral hydration in the asthenosphere. Also, water cannot cause the viscosity minimum zone in the upper mantle. And finally, the dominant mechanism responsible for hotspot immobility cannot be water content differences between their source and surrounding regions.

Diffusion creep and dislocation creep are two important mechanisms that dominate the plastic deformation of rocks and minerals in Earth's interior. Experimental deformation studies have suggested that incorporation of water in olivine significantly enhances both dislocation and diffusion creep rates^{1–4,7}. However, we note that those studies used polycrystalline olivine samples with over-saturated water. In such samples, large amounts of free water may have existed on grain boundaries, leading to a large enhancement of grain boundary sliding (or pressure-solution-accommodated creep), rather than dislocation creep or diffusion creep in the grain interior. On the other hand, the upper mantle is water unsaturated and free water is unlikely to exist. Therefore, the enhancement of creep rates by free water cannot occur in the real upper mantle. We also note that the ranges of water contents ($C_{\text{H}_2\text{O}} < 80$ wt p.p.m.) in these deformation studies^{1–4,7} are too narrow to determine accurately the effect of water on stress-strain rate measurements. This can lead to large errors in estimating the effect of water on mantle rheology.

Another problem with such rock deformation experiments is the very high stress (typically a hundred times higher than that in Earth's interior) needed to obtain experimentally determinable strain rates. High stress causes anomalously high-density dislocations, stacking faults and sub-grain boundaries, which may lead to artificial results for the Earth's interior. Instead, the measurement of self-diffusion coefficients in minerals is an independent way to study mantle rheology because high-temperature mineral creep is believed to be controlled by self-diffusion of the slowest species^{5,6} (which is silicon in the case of olivine^{8,9}). It allows a much wider range of experimental

conditions (such as pressure and $C_{\text{H}_2\text{O}}$) and also does not induce unrealistically high defect densities.

Costa and Chakraborty⁸ measured silicon self-diffusion coefficients (D_{Si}) in olivine single crystals with $C_{\text{H}_2\text{O}}$ values of ~ 40 and 370 parts per million by weight (wt p.p.m.) and concluded that even 45 wt p.p.m. of water enhances D_{Si} by two to three orders of magnitude by comparison with the results obtained under dry conditions by ref. 10. However, the data of ref. 8 did not show a systematic change in D_{Si} with $C_{\text{H}_2\text{O}}$ at ~ 40 and at ~ 370 wt p.p.m. In addition, our previous study¹¹ showed that ref. 10 may have underestimated D_{Si} under dry conditions. We therefore propose that the water effect was overestimated in ref. 8.

Here we systematically measured D_{Si} in olivine as a function of $C_{\text{H}_2\text{O}}$. Because the effects of iron on D_{Si} and on creep rates are very small under upper-mantle conditions^{11,12}, a single-crystal forsterite sample was used. We measured its D_{Si} at 8 GPa, 1,600 K and 1,800 K, and with well controlled $C_{\text{H}_2\text{O}}$ from < 1 up to about 800 wt p.p.m., which is realistic for the oceanic mantle. The experimental details are given in the Methods section.

Experimental results are shown in Fig. 1. D_{Si} systematically increases with increasing $C_{\text{H}_2\text{O}}$. D_{Si} values under wet conditions ($C_{\text{H}_2\text{O}} > 1$ wt p.p.m.) were fitted to the Arrhenius equation:

$$D_{\text{Si}} = A_0 C_{\text{H}_2\text{O}}^r \exp\left(-\frac{\Delta H}{RT}\right) \quad (1)$$

where A_0 is the pre-exponential factor, r is the $C_{\text{H}_2\text{O}}$ exponent, R is the gas constant, T is the absolute temperature, and ΔH is the activation

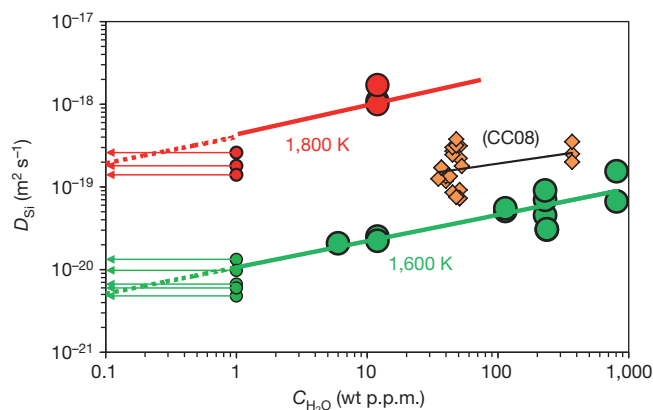


Figure 1 | D_{Si} versus $C_{\text{H}_2\text{O}}$ at 1,600 K and 1,800 K. The data points shown by small circles with an arrow are taken from ref. 11 on D_{Si} in dry forsterite at 8 GPa, with $C_{\text{H}_2\text{O}} < 1$ wt p.p.m.; these are below the detection resolution of FT-IR and SIMS. It was impossible to obtain data points at 1,800 K with high $C_{\text{H}_2\text{O}}$ because of the low melting temperature of hydrous forsterite³¹. Even when $C_{\text{H}_2\text{O}}$ was low, the isotopically enriched thin-film coating of the diffusion couple was often damaged during annealing at this temperature. CC08 indicates data points (orange diamonds) taken from Costa and Chakraborty⁸, normalized to 1,600 K and 8 GPa, using an activation energy of 358 kJ mol^{-1} from ref. 8 and an activation volume of $1.7 \text{ cm}^3 \text{ mol}^{-1}$ from ref. 11.

¹Bayerisches Geoinstitut, Universität Bayreuth, Bayreuth D95440, Germany. ²Helmholtz Centre Potsdam, Potsdam D14473, Germany. ³Institute for Study of the Earth's Interior, Okayama University, Misasa, Tottori 682-0193, Japan.

enthalpy. By fitting the experimental results to equation (1), we determined A_0 , r and ΔH to be $10^{-5.8 \pm 0.7} \text{ m}^2 \text{ s}^{-1}$, 0.32 ± 0.07 and $434 \pm 20 \text{ kJ mol}^{-1}$, respectively. The activation energy ΔE is $420 \pm 23 \text{ kJ mol}^{-1}$ (after a pressure correction using an activation volume of $1.7 \pm 0.4 \text{ cm}^3 \text{ mol}^{-1}$; ref. 11), which is essentially the same as that for dry conditions ($410 \pm 30 \text{ kJ mol}^{-1}$; see Supplementary Information)¹¹.

The present results thus demonstrate $D_{\text{Si}} \propto (\text{C}_{\text{H}_2\text{O}})^{0.32 \pm 0.07} \approx (\text{C}_{\text{H}_2\text{O}})^{1/3}$. Given that $[\text{V}_{\text{Si}}^{''''}]$ (where four primes indicates a charge of minus four on the Si vacancies) is proportional to $(\text{C}_{\text{H}_2\text{O}})^{2/3}$ (ref. 13) under the charge-neutral conditions (where the dot indicates one positive charge on the hydroxide ion on the oxygen site), $[(\text{OH})_{\text{O}}^{\bullet}] = 2[\text{V}_{\text{Mg}}^{''}]$ (ref. 14), a special explanation is necessary for the $\text{C}_{\text{H}_2\text{O}}$ exponent of D_{Si} . One hypothesis accounting for the $\text{C}_{\text{H}_2\text{O}}$ exponent is that Si diffusion is controlled by $\text{V}_{\text{O}}^{\bullet\bullet}$ as well as $\text{V}_{\text{Si}}^{''''}$. The Si^{4+} in forsterite is tightly surrounded by O^{2-} in a tetrahedron. If an oxygen ion is missing, the hopping probability of $\text{V}_{\text{Si}}^{''''}$ should greatly increase. Hence, Si diffusion may be dominated by $\text{V}_{\text{O}}^{\bullet\bullet}$ -associated $\text{V}_{\text{Si}}^{''''}$. Although the $[\text{V}_{\text{O}}^{\bullet\bullet}]$ is low, a certain proportion of $\text{V}_{\text{Si}}^{''''}$ should be associated with $\text{V}_{\text{O}}^{\bullet\bullet}$ owing to the Coulomb potential. As a result, D_{Si} should be proportional to both $[\text{V}_{\text{Si}}^{''''}]$ and $[\text{V}_{\text{O}}^{\bullet\bullet}]$. Given that $[\text{V}_{\text{O}}^{\bullet\bullet}] \propto (\text{C}_{\text{H}_2\text{O}})^{-1/3}$ (ref. 13), we have $D_{\text{Si}} \propto [\text{V}_{\text{Si}}^{''''}] \times [\text{V}_{\text{O}}^{\bullet\bullet}] \propto (\text{C}_{\text{H}_2\text{O}})^{2/3} \times (\text{C}_{\text{H}_2\text{O}})^{-1/3} = (\text{C}_{\text{H}_2\text{O}})^{1/3}$. However, we do not know what proportion of $\text{V}_{\text{Si}}^{''''}$ is associated with $\text{V}_{\text{O}}^{\bullet\bullet}$. It is possible that all the $\text{V}_{\text{Si}}^{''''}$ are associated with $\text{V}_{\text{O}}^{\bullet\bullet}$ because of the high Coulomb potential. In this case, D_{Si} would not be proportional to $[\text{V}_{\text{O}}^{\bullet\bullet}]$. Further investigation is required to explain the observed small $\text{C}_{\text{H}_2\text{O}}$ exponent of D_{Si} in view of defect chemistry.

Natural iron-bearing olivine in the real mantle can contain small amounts of $\text{Fe}_{\text{Mg}}^{\bullet}$, (where Fe^{3+} on a Fe or Mg metal site has an excess charge of +1) which may change the charge neutrality conditions and the $\text{C}_{\text{H}_2\text{O}}$ exponents for D_{Si} . However, the D_{Si} obtained in natural olivine by ref. 8 under wet conditions at high pressure showed essentially the same increase with increasing $\text{C}_{\text{H}_2\text{O}}$ from 30–50 wt p.p.m. to 370 wt p.p.m., as shown in Fig. 1. It suggests that $\text{Fe}_{\text{Mg}}^{\bullet}$ in natural olivine is not essential for Si diffusion in the investigated $\text{C}_{\text{H}_2\text{O}}$ range.

In the case that $\text{C}_{\text{H}_2\text{O}}$ is extremely high, the defect chemistry could be changed by incorporation of protons in Si vacancies and the hydrated Si vacancies— $\text{H}_{\text{Si}}^{''}$, $(2\text{H})_{\text{Si}}^{''}$, $(3\text{H})_{\text{Si}}^{''}$ and $(4\text{H})_{\text{Si}}^{''}$ (where the superscript cross indicates no excess charge on the Si vacancy)—whose concentrations have a larger $\text{C}_{\text{H}_2\text{O}}$ exponent (that is, 0.5–2) than $\text{V}_{\text{Si}}^{''''}$ (ref. 13), could dominate Si diffusion, possibly leading to a hydrolytic weakening of olivine¹⁵. For this reason, we expect D_{Si} to have a larger $\text{C}_{\text{H}_2\text{O}}$ exponent under high $\text{C}_{\text{H}_2\text{O}}$ conditions. However, our experimental results do not show an increase in the $\text{C}_{\text{H}_2\text{O}}$ exponent up to 800 wt p.p.m. Higher $\text{C}_{\text{H}_2\text{O}}$ conditions are unlikely in the upper mantle except for in the mantle wedge, judging from petrological studies (~ 70 – 160 wt p.p.m. of water in depleted mantle¹⁶, and a value four to five times higher in enriched mantle^{17,18}). Therefore, the $\text{C}_{\text{H}_2\text{O}}$ exponent of 1/3 is the maximum for the majority of the upper mantle.

Diffusion creep and dislocation creep in olivine under high temperatures are thought to be controlled by Si self-diffusion^{5,6}. Therefore, the $\text{C}_{\text{H}_2\text{O}}$ exponent for D_{Si} should be identical to that for creep rates. However, deformation studies^{1–3} on olivine aggregates claimed a much larger $\text{C}_{\text{H}_2\text{O}}$ exponent, 1.2 ± 0.4 (Fig. 2). We found that the infrared spectra in these studies^{1–3} showed tiny sharp peaks with a broad band. This suggests that most of the water existed on grain boundaries. Accordingly, the high strain rates in their wet samples might have been caused by grain boundary sliding enhanced by free water on grain boundaries (see further discussion in the Supplementary Information). This idea is also supported by the much lower creep rates obtained in single crystals of hydrous olivine¹⁹ than those in polycrystalline (Fig. 2). Free water is unlikely to be present in the upper mantle (except for the mantle wedge), owing to the water-unsaturated conditions. In addition, the grain size is on the order of millimetres to centimetres in the upper mantle²⁰, meaning grain boundary sliding

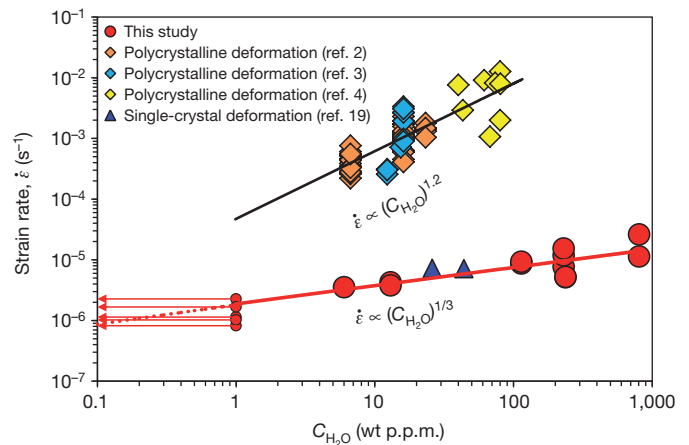


Figure 2 | Strain rate versus $\text{C}_{\text{H}_2\text{O}}$. D_{Si} values from this study are converted to strain rate using the proportional relationship of D_{Si} and strain rate⁶ with parameters from ref. 13. All data are normalized to a pressure of 8 GPa, a temperature of 1,600 K, and a stress of 300 MPa using an activation volume of $1.7 \text{ cm}^3 \text{ mol}^{-1}$ (ref. 11), activation energy of 420 kJ mol^{-1} , and a stress exponent of 3.5. The data points for $\text{C}_{\text{H}_2\text{O}} < 1$ wt p.p.m. are treated in the same way as in Fig. 1.

would be negligible⁷. Therefore, the creep rates of minerals in the real mantle cannot be enhanced by free water on grain boundaries.

Based on the small $\text{C}_{\text{H}_2\text{O}}$ exponent ($r = 1/3$) determined in this study, the difference in D_{Si} , as well as creep rates, between rheologically dry (< 1 wt p.p.m.) and maximum $\text{C}_{\text{H}_2\text{O}}$ of olivine in upper mantle ($< 1,000$ wt p.p.m.; refs 16–18) is within one order of magnitude. Because the variance of $\text{C}_{\text{H}_2\text{O}}$ in the upper mantle is very small, that is, ~ 100 – $1,000$ wt p.p.m. (refs 16–18), such a small range only causes ~ 0.3 orders of magnitude difference in creep rates. This is much smaller than other factors that affect rheological properties like temperature or shear stress. Hence, we conclude that the effect of water on upper mantle rheology is not significant, which is in complete contrast to what has been commonly accepted to be the case^{7,8,21,22}.

This small effect of water on upper-mantle rheology means that many geodynamical problems must be reconsidered. Two ideas, partial melting and hydration^{7,21–23}, have been commonly considered to explain plate motion because both could soften the oceanic asthenosphere. Previous overestimates of water effects on creep rates have erroneously supported the idea that hydration is the main reason for plate motion^{7,21,22}. Using the $\text{C}_{\text{H}_2\text{O}}$ exponent of 1/3, if 75% of the original water is extracted during mantle dehydration (~ 110 wt p.p.m. of water before dehydration¹⁶, and ~ 28 wt p.p.m. after dehydration²⁴), the creep rates change only by a factor of 1.6. On the other hand, the melt fraction in the asthenosphere is estimated to be 1.25–0.25% (ref. 25) or less²⁶. Such a small melt fraction enhances the creep rates by at most a factor of three²³. However, the high geothermal gradient in the oceanic mantle at < 200 km, and especially at < 100 km (about 12 K km^{-1})²⁷, causes the creep rates to increase by at least six orders of magnitude from a depth of 60 km to a depth of 200 km. Thus, the effect of temperature gradient on creep rates appears to be much larger than that of $\text{C}_{\text{H}_2\text{O}}$ or melt fraction. The softening of the oceanic asthenosphere that allows plate motion cannot occur by hydration or by partial melting.

In addition, the presence of a minimum-viscosity zone has been expected in the asthenosphere based on the seismically observed low-velocity and high-attenuation zone²⁸. However, because the effect of pressure on D_{Si} is also small¹¹, the viscosity in the upper mantle (which is calculated using the inverse relationship between D_{Si} and viscosity²⁹ based on oceanic geotherm²⁷) decreases monotonically with increasing depth (Fig. 3) even if the geothermal gradient is very small (that is, $< 1 \text{ K km}^{-1}$) at a depth exceeding 200 km. Thus, on the basis of the values of D_{Si} and taking the effects of pressure, temperature, and

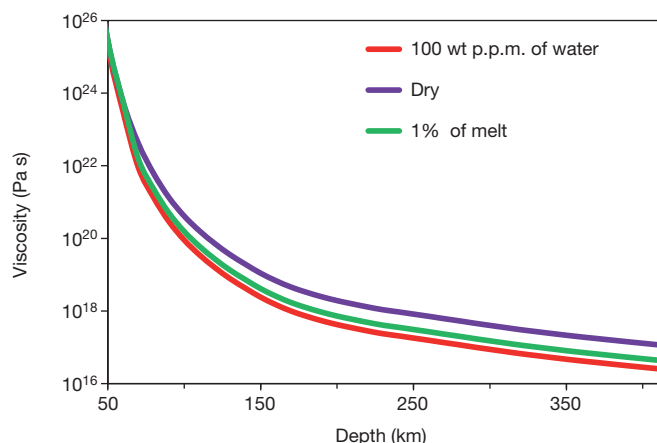


Figure 3 | Viscosity in the upper mantle. Viscosity η is calculated from D_{Si} using the inverse relationship of η and D_{Si} (ref. 29), that is, $\eta = 10kTr_c^2/(D_{\text{Si}}m_a)$, where k is the Boltzmann constant, T is the absolute temperature based on the oceanic geotherm²⁷, r_c is the crystal radius, and m_a is the mass of a Si ion. The grain size in the mantle is assumed to be ~ 1 mm. D_{Si} is a function of temperature, $C_{\text{H}_2\text{O}}$ and pressure, as given by equation (1) with $\Delta H = \Delta E + P\Delta V$, for which activation energy ΔE and activation volume ΔV values of 420 kJ mol^{-1} and $1.7 \text{ cm}^3 \text{ mol}^{-1}$ were used¹¹, respectively. The influence of partial melting on viscosity is calculated from the melt fraction dependence of creep rates²³.

water content into account, the minimum-viscosity zone does not appear in the asthenosphere.

Finally, an open question in mantle dynamics is why hotspots are so immobile in the face of plate motion. If it were true that water has a large effect on mantle rheology, high values of $C_{\text{H}_2\text{O}}$ in the source regions of hotspots in comparison to that in surrounding regions would be a possible explanation. However, our results demonstrate that this idea is not valid. Taking the Hawaii hotspot as an example, the $C_{\text{H}_2\text{O}}$ in its source is $\sim 750 \text{ wt p.p.m.}$, and $\sim 110 \text{ wt p.p.m.}$ in the surrounding regions¹⁷. Our results indicate that this difference would cause a viscosity contrast of a factor of two, which is rather small in comparison with that caused by temperature difference ($\sim 200 \text{ K}$ hotter than surrounding mantle³⁰, resulting in a viscosity decrease by more than one order of magnitude). Hence, the $C_{\text{H}_2\text{O}}$ contrast cannot be the major reason for the immobility of hotspots.

METHODS SUMMARY

A synthetically produced forsterite single crystal was cored into disks with diameter 1 mm and thickness 1 mm, which were used as the starting materials. The chemical composition of the crystal was Mg_2SiO_4 with Ir the major impurity ($\sim 80 \text{ wt p.p.m.}$). The cored disks were doped with water at 8 GPa and 1,600 K using talc + brucite as a water source with enstatite + graphite/gold powder, using a multi-anvil apparatus. The variation of water contents in the samples were made by varying the ratio of the water source to the enstatite + graphite/gold powder. After being carefully polished in an alkaline colloidal silica solution, each water-doped disk was coated with a $\sim 500 \text{ nm}$ ^{29}Si -enriched Mg_2SiO_4 thin film by a pulsed laser deposition system, and then annealed at 8 GPa and 1,600 K or 1,800 K for diffusion. The water contents in the samples were determined using Fourier transform infrared (FT-IR) spectroscopy both before and after diffusion annealing. The sample surfaces with the thin film were polished again in the alkaline solution after diffusion annealing to reduce their surface roughness, which could lead to large analytical uncertainties in diffusion profile analyses. The diffusion profiles were obtained using a Cameca 6f secondary ion mass spectrometer (SIMS). The diffusion coefficients were calculated by fitting the profiles to the solution of Fick's second law.

Full Methods and any associated references are available in the online version of the paper.

Received 31 August 2012; accepted 12 April 2013.

1. Mei, S. & Kohlstedt, D. L. Influence of water on plastic deformation of olivine aggregates. 1. Diffusion creep regime. *J. Geophys. Res.* **105**, 21457–21469 (2000).

2. Mei, S. & Kohlstedt, D. L. Influence of water on plastic deformation of olivine aggregates. 2. Dislocation creep regime. *J. Geophys. Res.* **105**, 21471–21481 (2000).
3. Karato, S. I., Paterson, M. S. & Fitzgerald, J. D. Rheology of synthetic olivine aggregates: influence of grain size and water. *J. Geophys. Res.* **91**, 8151–8176 (1986).
4. Jung, H. & Karato, S. Water-induced fabric transitions in olivine. *Science* **293**, 1460–1463 (2001).
5. Frost, H. J. & Ashby, M. F. *Deformation Mechanism Maps* Ch. 2 6–16 (Pergamon Press, 1982).
6. Weertman, J. in *Mechanics and Materials: Fundamentals and Linkages* (eds Meyers, M. A., Armstrong, R. W. & Kirschner, H.) 451–488 (Wiley, 1999).
7. Hirth, G. & Kohlstedt, D. L. Rheology of the upper mantle and the mantle wedge: a view from the experimentalists. *Geophys. Monogr. Ser.* **138**, 83–105 (2003).
8. Costa, F. & Chakraborty, S. The effect of water on Si and O diffusion rates in olivine and implications for transport properties and processes in the upper mantle. *Phys. Earth Planet. Inter.* **166**, 11–29 (2008).
9. Houlier, B., Cheraghmakani, M. & Jaoul, O. Silicon diffusion in San Carlos olivine. *Phys. Earth Planet. Inter.* **62**, 329–340 (1990).
10. Dohmen, R., Chakraborty, S. & Becker, H. W. Si and O diffusion in olivine and implications for characterizing plastic flow in the mantle. *Geophys. Res. Lett.* **29**, 2030, <http://dx.doi.org/10.1029/2002GL015480> (2002).
11. Fei, H. et al. High silicon self-diffusion coefficient in dry forsterite. *Earth Planet. Sci. Lett.* **345–348**, 95–103 (2012).
12. Durham, W. B. & Goetze, C. A comparison of the creep properties of pure forsterite and iron-bearing olivine. *Tectonophysics* **40**, T15–T18 (1977).
13. Kohlstedt, D. L. The role of water in high-temperature rock deformation. *Rev. Mineral. Geochem.* **62**, 377–396 (2006).
14. Kohlstedt, D. L., Keppler, H. & Rubie, D. C. Solubility of water in the α , β and γ phases of $(\text{Mg}, \text{Fe})_2\text{SiO}_4$. *Contrib. Mineral. Petrol.* **123**, 345–357 (1996).
15. Brodholt, J. P. & Refson, K. An ab initio study of hydrogen in forsterite and a possible mechanism for hydrolytic weakening. *J. Geophys. Res.* **105**, 18977–18982 (2000).
16. Workman, R. K. & Hart, S. R. Major and trace element composition of the depleted MORB mantle (DMM). *Earth Planet. Sci. Lett.* **231**, 53–72 (2005).
17. Dixon, J. E., Leist, L., Langmuir, C. & Schilling, J. G. Recycled dehydrated lithosphere observed in plume-influenced mid-ocean-ridge basalt. *Nature* **420**, 385–389 (2002).
18. Hirschmann, M. M. Water, melting, and the deep Earth H_2O cycle. *Annu. Rev. Earth Planet. Sci.* **34**, 629–653 (2006).
19. Ratteron, P. et al. Experimental deformation of olivine single crystals at mantle pressures and temperatures. *Phys. Earth Planet. Inter.* **172**, 74–83 (2009).
20. Karato, S. I. Grain-size distribution and rheology of the upper mantle. *Tectonophysics* **104**, 155–176 (1984).
21. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
22. Karato, S. I. & Jung, H. Water, partial melting and the origin of the seismic low velocity and high attenuation zone in the upper mantle. *Earth Planet. Sci. Lett.* **157**, 193–207 (1998).
23. Mei, S., Bai, W., Hiraga, T. & Kohlstedt, D. L. Influence of melt on the creep behavior of olivine–basalt aggregates under hydrous conditions. *Earth Planet. Sci. Lett.* **201**, 491–507 (2002).
24. Bell, D. R. & Rossman, G. R. Water in Earth's mantle: the role of nominally anhydrous minerals. *Science* **255**, 1391–1397 (1992).
25. Kawakatsu, H. et al. Seismic evidence for sharp lithosphere–asthenosphere boundaries of oceanic plates. *Science* **324**, 499–502 (2009).
26. Hirschmann, M. M. Partial melt in the oceanic low velocity zone. *Phys. Earth Planet. Inter.* **179**, 60–71 (2010).
27. Green, D. H. & Ringwood, A. E. The genesis of basaltic magmas. *Contrib. Mineral. Petrol.* **15**, 103–190 (1967).
28. Anderson, D. L. Earth's viscosity. *Science* **151**, 321–322 (1966).
29. McKenzie, D. P. The viscosity of the mantle. *Geophys. J. R. Astron. Soc.* **14**, 297–305 (1967).
30. Putirka, K. D. Mantle potential temperatures at Hawaii, Iceland, and the mid-ocean ridge system, as inferred from olivine phenocrysts: evidence for thermally driven mantle plumes. *Geochim. Geophys. Geosyst.* **6**, Q05L08 (2005).
31. Inoue, T. Effect of water on melting phase relations and melt composition in the system $\text{Mg}_2\text{SiO}_4\text{--MgSiO}_3\text{--H}_2\text{O}$ up to 15 GPa. *Phys. Earth Planet. Inter.* **85**, 237–263 (1994).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Chakraborty and R. Dohmen at Ruhr-University of Bochum for thin-film deposition and discussions about experimental methods. We also thank A. Yoneda at Okayama University for providing the single crystal, H. Keppler for FT-IR measurement, A. Audétat for ICP-MS analysis, and T. Boffa-Ballaran for X-ray diffraction analysis. We acknowledge support from the ENB (Elite Network Bavaria) programmes.

Author Contributions T.K. organized the project. The samples were prepared by H.F. and D.Y. All high pressure experiments and FT-IR measurements were performed by H.F. SIMS analyses were made by H.F. and M.W. The manuscript was completed by H.F. and T.K.; all authors read and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.F. (hongzhan.fei@uni-bayreuth.de).

METHODS

Starting material. A single-crystal forsterite sample was obtained from Oxide Company, Japan. The chemical composition of the crystal is Mg_2SiO_4 . Its trace-element compositions were obtained from ref. 11. No O–H absorption bands were detected by Fourier transform infrared (FT-IR), indicating that the water content was less than 1 wt.p.p.m. We used disks cored from the crystal, with diameter 1 mm and thickness 1 mm and the thickness oriented along the *b* axis.

Water-doping experiments. The cored forsterite disks were pre-annealed at 8 GPa and 1,600 K in the presence of a water source. This step is necessary to equilibrate the water in the crystal before diffusion annealing.

Each forsterite disk was loaded into a platinum capsule, with an outer diameter of 2.0 mm and an inner diameter of 1.6 mm, with one end sealed. A mixture of talc and brucite powders (weight ratio 4:1) was used as the water source and also to control the silica activity in the capsule. The space between the forsterite disk and capsule wall was filled with graphite or gold + enstatite (weight ratio 35:1) powder for low- and high-water-content experiments, respectively, to protect the single crystal from mechanical damage at high pressure (Supplementary Fig. 1). The capsule was closed and sealed by arc welding in liquid nitrogen to minimize water escape from the capsule. The water content in the capsule was controlled by the ratio of water source to graphite or gold + enstatite. In dry experiments, graphite powder was loaded around the samples; the capsules were then dried in a vacuum oven at 470 K for at least 24 h and sealed on a hot plate to minimize the amount of moisture absorbed from the atmosphere. The final length of capsules was 4–4.5 mm.

High-pressure experiments were performed using a Kawai-type multi-anvil apparatus at the University of Bayreuth. All experiments were performed at 8 GPa and 1,600 K. In each run, the sealed platinum capsule was located in a MgO cylinder in a LaCrO₃ stepped heater with a ZrO₂ thermal insulator. A MgO octahedron (with 5 wt% Cr₂O₃) with edge length 14 mm was used as the pressure medium (Supplementary Fig. 1). Eight tungsten carbide cubes with 32-mm edge length and 8-mm truncation edge length were used to generate high pressures. The temperatures were measured using a W97%Re3%-W75%Re25% thermocouple, 0.25 mm in diameter, whose junction was placed at the bottom of the capsule. The assembly was compressed to the target pressure over 2–4 h, heated to 1,273 K at a rate of 50 K min^{−1}, kept at 1,273 K for 1 h to decompose talc and brucite and to make the water distribution homogenous in the capsule, the assembly was then heated to 1,600 K in 5 min and kept for a long duration for water equilibration (50–70 h), as calculated from the hydrogen diffusion coefficients in forsterite³². The temperature was under automatic control, thus limiting variation to less than 2 K during annealing. After annealing, the sample was quenched by switching off the heating power and gradually decompressed to ambient pressure over a long period (15–20 h) to prevent crystal breakage.

The forsterite disks were recovered by cutting into the platinum capsule using a steel blade. No obvious cracks were found in the samples if small amounts of water source were used. With high amounts of water source, the crystal always contained some cracks and broke into pieces. However, in such cases we were still able to find usable pieces for diffusion experiments.

Deposition. The water-doped samples were polished using diamond powders with grain sizes of 0.25 µm, followed by an alkaline colloidal silica solution for >3 h until all small scratches were removed. The highly polished surface was then coated with ~500 nm of ²⁹Si enriched Mg₂SiO₄ and 100 nm of ZrO₂ using a pulsed laser deposition system at the Ruhr-University of Bochum³³. We also conducted some diffusion experiments without the ZrO₂ film for comparison, and showed that the ZrO₂ does not affect *D*_{Si}, which was already confirmed in our previous study¹¹. Prior to each deposition, the samples were heated up to 470 K for 10–15 min in the vacuum chamber of the pulsed laser deposition system so as to remove any free water from the sample surface. The structural water in the crystals did not escape during this step.

Diffusion annealing. Each thin-film-coated sample was placed in a platinum capsule with the same ratio of water source and graphite or gold + enstatite as

used for the corresponding water-doping experiment and was then annealed at 8 GPa and 1,600 K or 1,800 K using the same high-pressure assembly (Supplementary Fig. 1). The annealing durations, ranging from 5–41 h as summarized in Supplementary Table 1, were estimated from silicon diffusion coefficient data for olivine⁸ and forsterite¹¹.

FT-IR analysis. The water contents in the samples after water-doping experiments and also after diffusion annealing were measured using a high-resolution FT-IR spectrometer at the University of Bayreuth, described in ref. 11. Each forsterite sample for FT-IR analysis was polished on both faces normal to the *b* axis using 0.25-µm diamond powder. Two hundred scans were accumulated for each spectrum at a resolution of 1 cm^{−1}. Two or three spectra were obtained for each sample with at least one near the centre of the disk and one near the edge. One sample (V720) was also polished parallel to the *b* axis, and the water content was obtained as a function of distance from the coated thin film at 60-µm steps. After a background baseline correction and thickness normalization to 1 cm, the water contents were determined using the calibration given by³⁴

$$C_{\text{H}_2\text{O}} = 0.188 \times \int k(\nu) d\nu \quad (2)$$

where *C*_{H₂O} was the water content in wt.p.p.m. and *k*(*ν*) was the absorption coefficient at wavenumber *ν*. Integration was performed between 3,000 cm^{−1} and 4,000 cm^{−1} (ref. 11). The results of *C*_{H₂O} in the samples are shown in the Supplementary Information.

SIMS analysis. The apparent diffusion profiles were measured by secondary ion mass spectrometry (SIMS) depth profiling using the Cameca IMS-6f installed at the Helmholtz Centre in Potsdam, Germany, with the same set-up for determining *D*_{Si} in dry forsterite as in our previous study¹¹. The depth of each SIMS crater was determined using a 3D-Nanofocus vertical microscope at the University of Bayreuth. The *D*_{Si} was obtained by fitting the data to the solution of Fick's second law

$$c = \frac{c_0 - c_1}{2} \operatorname{erf}\left(\frac{x-h}{\sqrt{4Dt + L^2(\sigma)}}\right) + \frac{c_0 + c_1}{2} \quad (3)$$

where *c* is the observed abundance of ²⁹Si, *c*₁ is the initial abundance of ²⁹Si in the isotopic film, *c*₀ is the initial abundance of ²⁹Si in the substrate, *x* is the distance from the surface, *h* is the position of the boundary between the thin film and substrate, *t* is the annealing time, *L*(*σ*) is the nominal diffusion length in zero-time diffusion runs related to surface roughness (discussed below), and erf(*z*) is the error function¹¹. An example of the diffusion profiles is shown in Supplementary Fig. 2.

Surface problem. Because of the crystallization of thin films, the surface roughness significantly increased during high-temperature annealing and became the major analytical uncertainty source¹¹. Hence, the sample surfaces after diffusion annealing were chemically polished in an alkaline colloidal silica solution until the roughness was reduced to <50 nm, measured with a 3D-Nanofocus vertical microscope at University of Bayreuth. Only a thin layer (<200 nm), located well beyond the apparent diffusion profile, was removed during the final chemical polishing¹¹. In addition, the apparent diffusion lengths obtained by SIMS were also corrected using a roughness calibration line obtained by a series of zero-time runs (equation (3)), in which the nominal diffusion lengths *L* are approximately a linear function of the standard deviation *σ* of the surface roughness at the bottoms of the craters (Supplementary Fig. 3). Detailed discussion about the surface problem is given in ref. 11.

32. Demouchy, S. & Mackwell, S. Water diffusion in synthetic iron-free forsterite. *Phys. Chem. Miner.* **30**, 486–494 (2003).
33. Dohmen, R. *et al.* Production of silicate thin films using pulsed laser deposition (PLD) and applications to studies in mineral kinetics. *Eur. J. Mineral.* **14**, 1155–1168 (2002).
34. Bell, D. R. *et al.* Hydroxide in olivine: a quantitative determination of the absolute amount and calibration of the IR spectrum. *J. Geophys. Res.* **108**(B2), 2105, <http://dx.doi.org/10.1029/2001JB000679> (2003).

Barium distributions in teeth reveal early-life dietary transitions in primates

Christine Austin^{1,2,3*}, Tanya M. Smith^{4*}, Asa Bradman⁵, Katie Hinde^{4,6}, Renaud Joannes-Boyau⁷, David Bishop⁸, Dominic J. Hare^{8,9}, Philip Doble⁸, Brenda Eskenazi⁵ & Manish Arora^{1,2,3}

Early-life dietary transitions reflect fundamental aspects of primate evolution and are important determinants of health in contemporary human populations^{1,2}. Weaning is critical to developmental and reproductive rates; early weaning can have detrimental health effects but enables shorter inter-birth intervals, which influences population growth³. Uncovering early-life dietary history in fossils is hampered by the absence of prospectively validated biomarkers that are not modified during fossilization⁴. Here we show that large dietary shifts in early life manifest as compositional variations in dental tissues. Teeth from human children and captive macaques, with prospectively recorded diet histories, demonstrate that barium (Ba) distributions accurately reflect dietary transitions from the introduction of mother's milk through the weaning process. We also document dietary transitions in a Middle Palaeolithic juvenile Neanderthal, which shows a pattern of exclusive breastfeeding for seven months, followed by seven months of supplementation. After this point, Ba levels in enamel returned to baseline prenatal levels, indicating an abrupt cessation of breastfeeding at 1.2 years of age. Integration of Ba spatial distributions and histological mapping of tooth formation enables novel studies of the evolution of human life history, dietary ontogeny in wild primates, and human health investigations through accurate reconstructions of breastfeeding history.

Weaning, the dietary transition from breast milk to exclusive solid food intake, concludes several years earlier in modern humans than in other great apes^{5,6}. Cross-cultural studies of nonindustrial societies reveal remarkable variation in weaning practices⁷. However, among non-human primates, dietary transitions remain understudied^{8,9}. In addition to the paucity of comparative primate data, our understanding of the evolution of human weaning has been limited by difficulties in assessing the precise timing and nature of dietary transitions during infancy⁷. Dental hard tissues are particularly valuable for reconstructing diet as they contain precise temporal and chemical records of early life⁴. Teeth begin forming *in utero*, record birth as the neonatal line, and manifest daily growth lines, which allow chronological ages to be determined at various positions within tooth crowns and roots (Supplementary Fig. 1).

We propose that micro-spatial analysis of barium/calcium ratios (Ba/Ca) in dental tissues represents a powerful approach to assess dietary transitions. Whereas prenatal Ba transfer is restricted by the placenta, marked enrichment occurs immediately after birth from mother's milk or infant formulas, which contain higher Ba levels than umbilical cord sera¹⁰. In response to these variations in dietary Ba exposure, Ba/Ca in enamel and dentine should increase at birth, remain elevated for the duration of exclusive breastfeeding and rise further with introduction of infant formula. Circulating Ba levels are

expected to change at weaning as Ba (and Ca) content and bioavailability is markedly different across plant and animal food sources^{11,12}. To test this hypothesis, we investigated Ba/Ca patterns in teeth from human children for whom early life diets were recorded prospectively, and in teeth from captive macaques in which maternal milk was collected and suckling behaviour observed.

High-resolution elemental analysis by laser ablation-inductively coupled plasma-mass spectrometry revealed marked Ba/Ca increases in enamel and dentine formed immediately after birth in human deciduous teeth ($n = 22$ of 25 individuals) (Fig. 1a–c). In 9 of 13 children who were initially breastfed and given infant formula later, two distinct zones of Ba/Ca distribution were apparent in postnatal regions formed before crown completion (Fig. 1d, e). Histological analysis (Supplementary Fig. 2) revealed a close correspondence between the formation time of the first zone and maternal reports of exclusive breastfeeding. Four individuals who continued to consume breast milk for a long period (9–42 months) after the introduction of formula at 1–2 months did not show two distinct Ba/Ca zones in enamel or dentine. In children for whom formula was introduced almost immediately after birth and who were breastfed for less than 1 month (Fig. 1e), the first Ba/Ca zone immediately adjacent to the neonatal line was narrower than in infants exclusively breastfed for longer. Individuals who were exclusively breastfed during the entire period of tooth crown formation ($n = 7$ of 25; Fig. 1f) showed an increase in Ba/Ca across the neonatal line, but as expected, no subsequent Ba/Ca zoning was apparent in postnatally formed dentine (as seen in infants who made a transition from breast milk to formula). Thus, Ba enrichment provides unambiguous evidence for postnatal feeding, as well as the beginning of supplementation; however, the transition from exclusive breastfeeding to formula intake may be obscured when breast milk remains the predominant dietary component after formula introduction. The extent of Ba/Ca increase at birth varies due to inter-individual differences in breast milk and formula Ba content. This is illustrated in Fig. 1f where the rise in Ba/Ca in response to breastfeeding is lower than in other individuals (Fig. 1d, e). Data on Ba/Ca values are given in Supplementary Table 1 and Supplementary Fig. 3.

Macaque permanent first molars also showed clear distinctions in Ba/Ca between pre- and postnatal regions, and close correspondence of postnatal changes in dental tissues and mother's milk (Supplementary Fig. 4). Although more diffuse owing to the nature of mineralization, Ba/Ca patterns in enamel correlated closely with dentine. Temporal mapping revealed Ba/Ca increases for the first 3–3.5 months of postnatal life (Fig. 2, Supplementary Figs 4–6 and Supplementary Table 2), followed by decreases that correlated with declines in suckling time and the initiation of solid food consumption. Moreover, Ba/Ca decreased more gradually during natural weaning than in

¹Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²Environmental and Occupational Medicine and Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ³Institute of Dental Research, Westmead Millennium Institute, Westmead Hospital, and Oral Pathology and Oral Medicine, Faculty of Dentistry, University of Sydney, Sydney, New South Wales 2145, Australia. ⁴Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁵Center for Environmental Research and Children's Health, School of Public Health, University of California, Berkeley, California 94720, USA. ⁶California National Primate Research Center, Davis, California 95616, USA. ⁷Southern Cross GeoScience, Southern Cross University, Lismore, New South Wales 2480, Australia. ⁸Elemental Bio-imaging Facility, University of Technology Sydney, Sydney, New South Wales 2007, Australia. ⁹The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Victoria 3010, Australia.

*These authors contributed equally to this work.

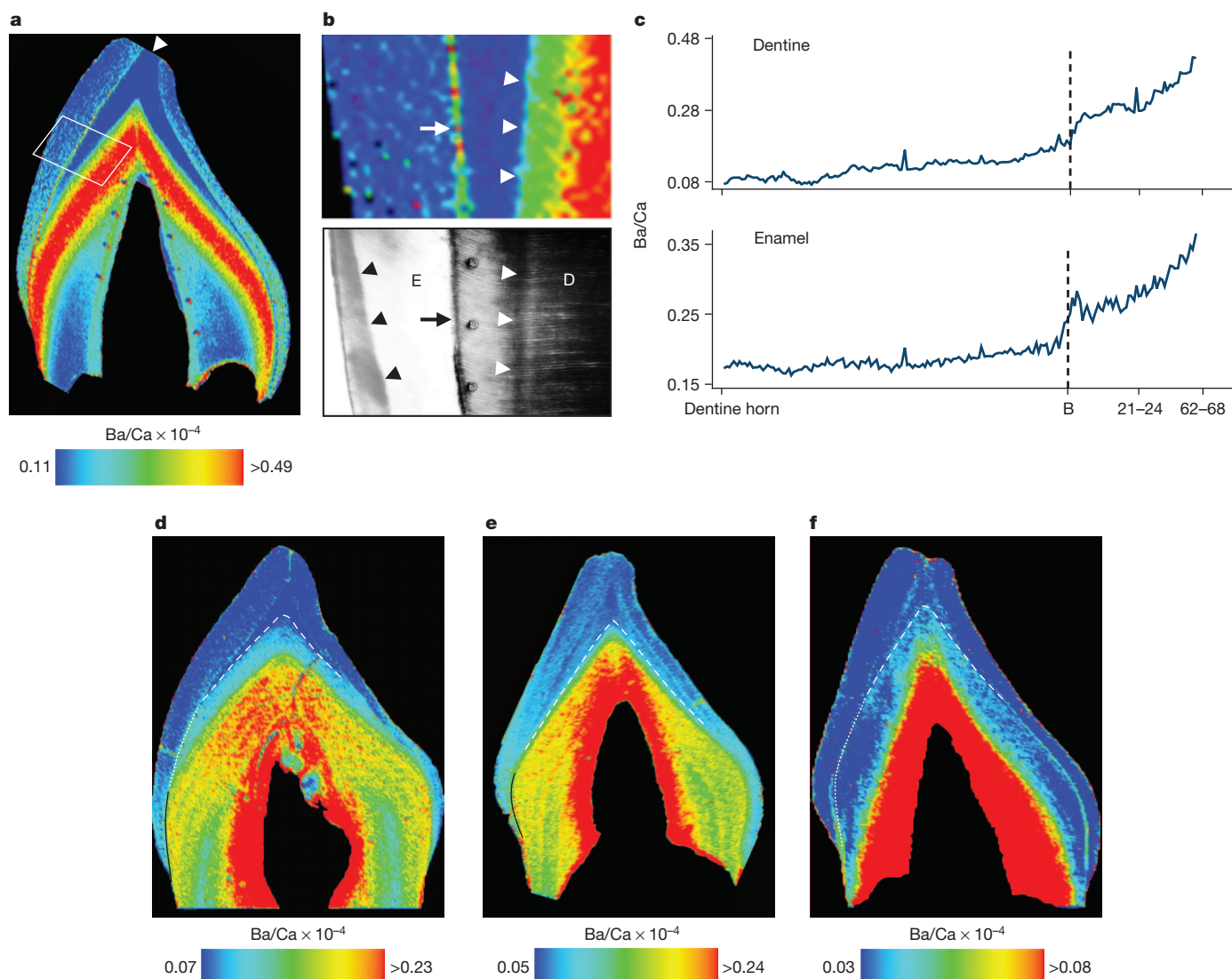


Figure 1 | Barium distribution in human deciduous teeth. **a**, Ba/Ca map of incisor. Dentine horn is indicated by an arrowhead. **b**, Area highlighted in **a** and polarized light micrograph. In dentine (D), Ba/Ca levels show a marked increase coinciding with the neonatal line (white arrowheads). The neonatal line in enamel (E) is indicated by black arrowheads and the enamel–dentine junction by arrows. **c**, Ba/Ca measured adjacent to the enamel–dentine junction from dentine horn to cervix of **a**, which rose at birth and with the introduction

of infant formula (21–24 days). The x axis shows days since birth (B). **d–f**, Three diet patterns: breastfeeding for 3 months (dotted white line) followed by exclusive formula feeding (solid black line) (**d**); formula introduced within 1 week of birth (solid black line) (**e**); and exclusive breastfeeding (dotted white line) (**f**). The neonatal line is indicated by a dashed white line. Intensity indices are $\text{Ba/Ca} \times 10^{-4}$. High Ba/Ca ratios adjacent to pulp (red zone) are in secondary dentine, a later-forming region not relevant to the current study (see ref. 12).

individuals who experienced truncated weaning periods (Fig. 2). In the most extreme case, an individual separated from its mother for several weeks at 166 days of age, precipitating cessation of milk synthesis and mammary gland involution, showed an abrupt Ba/Ca drop (Fig. 2c), which was independently estimated at 151–183 days of age.

Building upon our prospectively validated human and macaque results, we precisely documented diet transitions in a juvenile Neanderthal¹³. Barium is incorporated into the mineral phase (hydroxyapatite) during tooth calcification, which occurs rapidly after secretion in dentine, and more slowly and diffusely in enamel during maturation^{14,15}. Trace elements such as Ba are more resistant to post-mortem diagenetic alteration in enamel than in dentine, due in part to the greater original mineral content and lack of natural pores¹⁶. Thus, the distribution of Ba/Ca in well-preserved tooth enamel may yield direct information on early-life dietary transitions in fossil hominins.

Chemical and temporal mapping of Neanderthal first molar enamel (Fig. 3) revealed a transition pattern similar to the macaque that weaned abruptly. After approximately 13 days of prenatal enamel formation, Ba/Ca near the enamel–dentine junction increased and

remained elevated until approximately 227 days of age (~7.5 months), followed by intermediate values until 435 days of age (1.2 years). After this age Ba/Ca rapidly returned to prenatal levels for the final 1.15 years of crown formation. The Ba/Ca patterns in enamel were not observed in dentine due to diagenetic modification after death. However, diagenesis did not seem to have a significant influence on enamel, as concentrations of diagenetic indicators¹⁷ were low (Supplementary Table 3 and Supplementary Fig. 7). Furthermore, enamel Ba/Ca levels were similar to published values for other hominins^{11,18}, and Ba/Ca shifts were similar in form and timing between both mesial cusps, suggesting that the transition represents biogenic input rather than post-mortem modification (Supplementary Discussion). Although the subsurface occlusal and cervical enamel appears to show minor cracks that may lead to local modification¹⁷, most of the tooth crown is intact and naturally coloured. The *Sciadina* individual has also yielded mtDNA and enamel proteins^{19,20}, indicating that it is a well-preserved fossil.

Strontium/calcium ratios (Sr/Ca) in tooth enamel have been interpreted to reveal dietary transitions in baboons and humans^{9,21}. However, these events were inferred from species-typical norms or

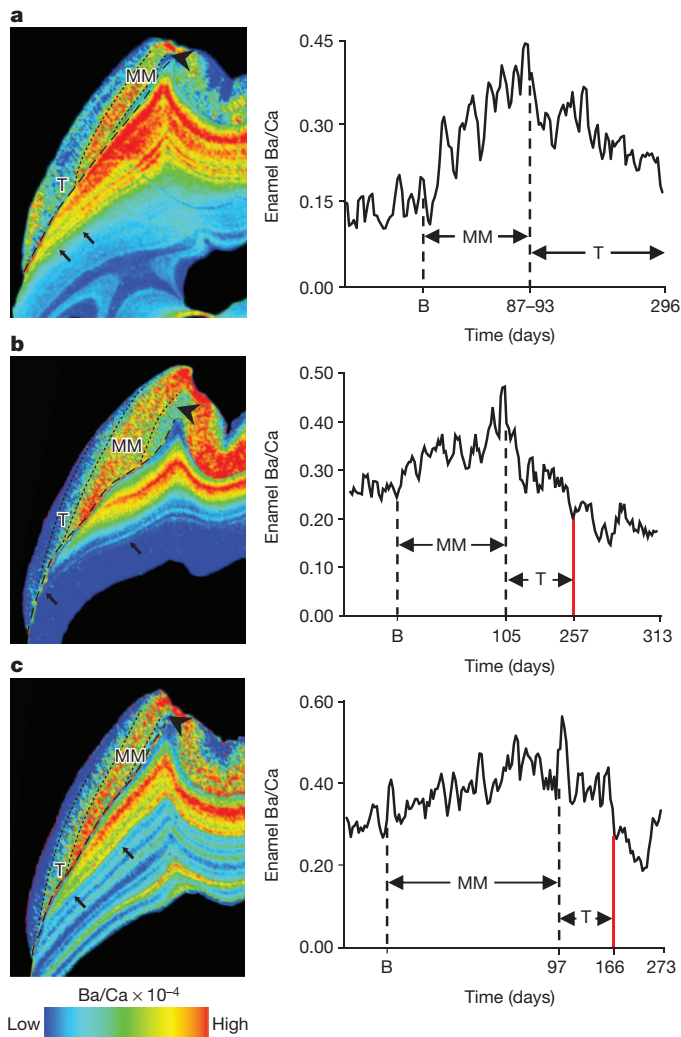


Figure 2 | Barium distribution reveals natural and truncated weaning.

a, Macaque 515: natural weaning after 296 days. **b**, Macaque 152: weaned slightly early due to maternal separation at 257 days. **c**, Macaque 401: markedly truncated weaning owing to maternal separation at 166 days. This individual's weight fluctuated during the final 7 months of life due to illness; post-weaning enrichment may be owing to release from skeletal stores³⁰. Diet transitions: prenatal regions (arrowhead), exclusive mother's milk (MM), transitional (T) periods, and post-weaning regions delineated in enamel (dotted lines) and dentine (black arrows). The enamel–dentine junction is indicated with a dashed line. The y axis shows enamel Ba/Ca adjacent to the enamel–dentine junction. The x axis shows days since birth (B) and weaning (red line). Elemental maps of dentine and enamel were rendered on different scales to show Ba/Ca transitions clearly.

recalled retrospectively years after the event, which may be subject to significant recall bias²². In light of this, and concerns that Sr might be more susceptible to diagenetic alteration than Ba due to its higher diffusivity^{23,24}, a posthoc comparison of Sr/Ca and Ba/Ca was conducted. We found that the reconstruction of diet history from Sr/Ca mapping was impeded due to proportionately smaller changes in Sr levels across transitions and inconsistent patterns in human and macaque samples (Supplementary Figs 3, 8 and 9, Supplementary Tables 4 and 5, and Supplementary Discussion). Two distinct regions between birth and 1.2 years were observed in the Neanderthal tooth for Ba/Ca and Sr/Ca, representing exclusive breastfeeding and solid food supplementation, although this is less clear from Sr/Ca when compared to Ba/Ca (Supplementary Fig. 10). Thus, Ba/Ca provides greater resolution of dietary transitions than Sr/Ca in extant and fossil material. Nonetheless, measurements of Sr isotopes in enamel have yielded useful data on diet and migration in early hominins¹⁸.

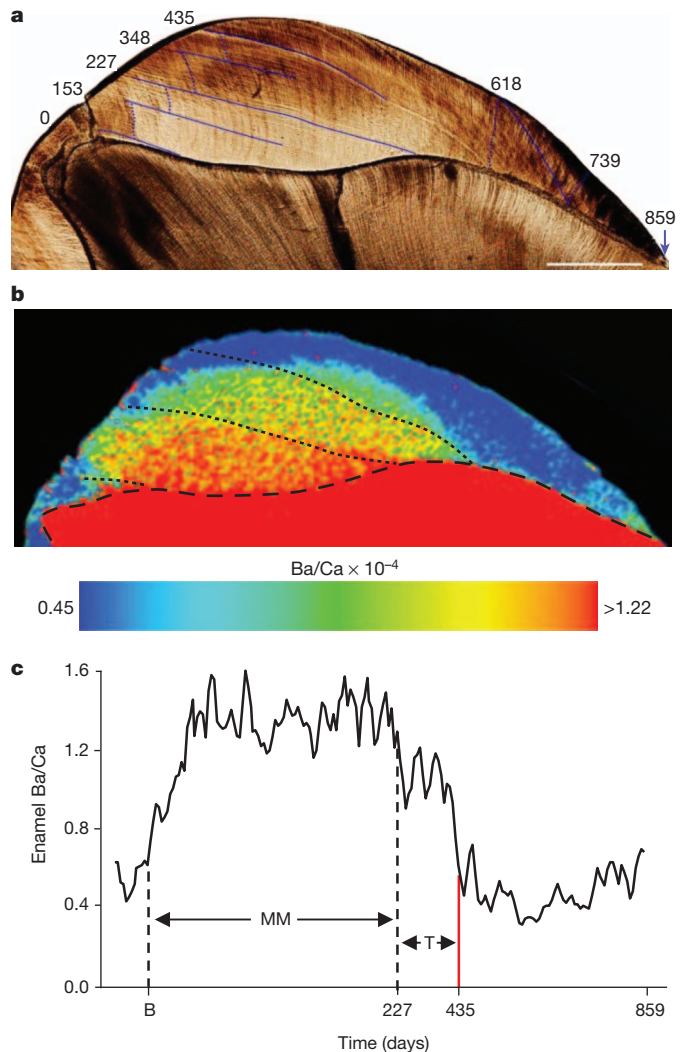


Figure 3 | Dietary transitions in a Neanderthal permanent first molar.

a, Developmental time (in days from birth) of stress lines in enamel (dark blue lines) was determined from daily growth increments (following dotted blue lines). Scale bar, 1 mm. **b**, Ba/Ca map shows marked variations in enamel at birth, 227 and 435 days, which resemble human and macaque transitions from exclusive maternal milk (MM) consumption to supplementation. **c**, Ba/Ca in enamel adjacent to the enamel–dentine junction. The x axis shows days from birth (B) to proposed exclusive MM, transitional diet (T) periods and hypothesized weaning event (red line). Elevated Ba/Ca levels at the very beginning and end of crown formation are probably due to subtle diagenetic modification¹⁷.

We have shown a direct correlation between Ba/Ca distributions in human deciduous teeth and breastfeeding data collected prospectively, thereby avoiding recall bias. In the macaques, patterns of suckling behaviour and Ba concentration in mother's milk are consistent with Ba/Ca in dental tissues, which consistently show a decrease in Ba/Ca from the onset of supplementation. Taken collectively, these results demonstrate that Ba/Ca in teeth effectively reflect Ba intake via mother's milk, and can be used to document developmental transitions in future studies of wild primate skeletal material, and for assessments of human health outcomes.

In the *Sceladina* Neanderthal, the protracted weaning process typical in primates was interrupted by unknown cause(s), precipitating abrupt cessation of suckling. The period of exclusive breastfeeding in this Neanderthal is consistent with other hominoids; human hunter-gatherers and wild chimpanzees also begin to supplement milk with solid food by around 6 months of age^{5,25}. Humans and chimpanzees may wean offspring as early as 1.0 and 4.2 years, respectively, without

serious health effects, but average 2.3–2.6 years⁵ and 5.3 years²⁵. When applied to additional samples, our approach will allow the evaluation of hypotheses that Neanderthal young routinely weaned at later ages than Upper Palaeolithic hominins²⁶, or possessed faster life histories than modern humans¹³, which have important implications for models of hominin population growth and species replacement.

METHODS SUMMARY

Human teeth were supplied from the Center for the Health Assessment of Mothers and Children of Salinas study, Monterey County, California, USA²⁷. Pregnant women were recruited before 20 weeks gestation, and data on breastfeeding and use of infant formulas were prospectively collected. From the 7-year assessment onwards, mothers were asked to bring in a tooth the child had shed, which was prepared according to standard histological techniques. The neonatal line was used to identify pre- and postnatal developmental periods. Prominent long-period incremental lines were mapped, and daily growth cross-striations in enamel were measured to determine the average daily enamel secretion rate. Macaque samples were obtained from two mother–infant dyads and two additional juveniles at the California National Primate Research Center (CNPRC), UC Davis, California, USA. Mothers and infants were captured for milk collection and morphometric measurements three times during lactation. Methods for rhesus macaque milk collection are described elsewhere⁸. In the week before milk collection, observations of infant suckling behaviour were recorded²⁸. Dentitions were collected opportunistically during animal necropsy in conjunction with the CNPRC Biological Specimens Program. First molars were dissected out after fixation, and histological sections were prepared and analysed following established protocols¹³. The Sladina Neanderthal upper first maxillary molar had been previously sectioned and temporally mapped¹³. Laser ablation–inductively coupled plasma–mass spectrometry (LA-ICP-MS) was used for elemental analysis of all samples according to published protocols²⁹. Instrument parameters were selected to generate images with pixel sizes of approximately 900 µm². Reported element ratios (Ba/Ca × 10^{−4} and Sr/Ca × 10^{−3}) were calculated from concentrations determined using ⁴³Ca, ⁸⁸Sr and ¹³⁸Ba with standard NIST 1486 bone meal. Other elements were quantified against NIST 612 as a standard. Changes in Ba/Ca were assigned ages by overlaying photomicrographs from histological temporal maps, which were registered along the enamel–dentine junction.

Full Methods and any associated references are available in the online version of the paper.

Received 3 December 2012; accepted 8 April 2013.

Published online 22 May 2013.

1. Ip, S. *et al.* Breastfeeding and maternal and infant health outcomes in developed countries. Evidence Report/Technology Assessment No. 153 (AHRQ, 2007).
2. McDade, T. W. Life history, maintenance, and the early origins of immune function. *Am. J. Hum. Biol.* **17**, 81–94 (2005).
3. Humphrey, L. T. Weaning behaviour in human evolution. *Semin. Cell Dev. Biol.* **21**, 453–461 (2010).
4. Smith, T. M. & Tafforeau, P. New visions of dental tissue research: Tooth development, chemistry, and structure. *Evol. Anthropol.* **17**, 213–226 (2008).
5. Sellen, D. W. & Smay, D. B. Relationship between subsistence and age at weaning in “preindustrial” societies. *Hum. Nature* **12**, 47–87 (2001).
6. Lee, P. C., Majluf, P. & Gordon, I. J. Growth, weaning and maternal investment from a comparative perspective. *J. Zool. (Lond.)* **225**, 99–114 (1991).
7. Sellen, D. W. Comparison of infant feeding patterns reported for nonindustrial populations with current recommendations. *J. Nutr.* **131**, 2707–2715 (2001).
8. Hinde, K., Power, M. L. & Oftedal, O. T. Rhesus macaque milk: Magnitude, sources, and consequences of individual variation over lactation. *Am. J. Phys. Anthropol.* **138**, 148–157 (2009).
9. Humphrey, L. T., Dirks, W., Dean, M. C. & Jeffries, T. E. Tracking dietary transitions in weanling baboons (*Papio hamadryas anubis*) using strontium/calcium ratios in enamel. *Folia Primatol. (Basel)* **79**, 197–212 (2008).
10. Krachler, M., Rossipal, E. & Micetic-turk, D. Concentrations of trace elements in sera of newborns, young infants, and adults. *Biol. Trace Elem. Res.* **68**, 121–135 (1999).
11. Sponheimer, M. & Lee-Thorp, J. A. Enamel diagenesis at South African Australopithecus sites: Implications for paleoecological reconstruction with trace elements. *Geochim. Cosmochim. Acta* **70**, 1644–1654 (2006).

12. Kohn, M. J., Morris, J. & Olin, P. Trace element concentrations in teeth – a modern Idaho baseline with implications for archaeometry, forensics, and palaeontology. *J. Archaeol. Sci.* **40**, 1689–1699 (2013).
13. Smith, T. M., Toussaint, M., Reid, D. J., Olejniczak, A. J. & Hublin, J.-J. Rapid dental development in a Middle Paleolithic Belgian Neanderthal. *Proc. Natl Acad. Sci. USA* **104**, 20220–20225 (2007).
14. Smith, C. E. Cellular and chemical events during enamel maturation. *Crit. Rev. Oral Biol. Med.* **9**, 128–161 (1998).
15. Goldberg, M., Kulkarni, A. B., Young, M. & Boskey, A. Dentin: structure, composition and mineralization. *Front. Biosci.* **3**, 711–735 (2011).
16. Kohn, M. J., Schoeninger, M. J. & Barker, W. W. Altered states: effects of diagenesis on fossil tooth chemistry. *Geochim. Cosmochim. Acta* **63**, 2737–2747 (1999).
17. Hinz, E. A. & Kohn, M. J. The effect of tissue structure and soil chemistry on trace element uptake in fossils. *Geochim. Cosmochim. Acta* **74**, 3213–3231 (2010).
18. Balter, V., Braga, J., Telouk, P. & Thackeray, J. F. Evidence for dietary change but not landscape use in South African early hominins. *Nature* **489**, 558–560 (2012).
19. Nielsen-Marsh, C. M. *et al.* Extraction and sequencing of human and Neanderthal mature enamel proteins using MALDI-TOF/TOF MS. *J. Archaeol. Sci.* **36**, 1758–1763 (2009).
20. Orlando, L. *et al.* Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr. Biol.* **16**, R400–R402 (2006).
21. Humphrey, L. T., Dean, M. C., Jeffries, T. E. & Penn, M. Unlocking evidence of early diet from tooth enamel. *Proc. Natl Acad. Sci. USA* **105**, 6834–6839 (2008).
22. Gillespie, B., d’Arcy, H., Schwartz, K., Bobo, J. & Foxman, B. Recall of age of weaning and other breastfeeding variables. *Int. Breastfeed. J.* **1**, 4 (2006).
23. Kohn, M. J. & Moses, R. J. Trace element diffusivities in bone rule out simple diffusive uptake during fossilization but explain *in vivo* uptake and release. *Proc. Natl Acad. Sci. USA* **110**, 419–424 (2013).
24. Ezzo, J. A. A test of diet versus diagenesis at Ventana Cave, Arizona. *J. Archaeol. Sci.* **19**, 23–37 (1992).
25. Smith, T. M. *et al.* First molar eruption, weaning, and life history in living wild chimpanzees. *Proc. Natl Acad. Sci. USA* **110**, 2787–2791 (2013).
26. Skinner, M. Dental wear in immature Late Pleistocene European hominines. *J. Archaeol. Sci.* **24**, 677–700 (1997).
27. Harley, K., Stamm, N. & Eskenazi, B. The Effect of time in the U.S. on the duration of breastfeeding in women of Mexican descent. *Matern. Child Health J.* **11**, 119–125 (2007).
28. Altmann, J. Observational study of behavior: Sampling methods. *Behaviour* **49**, 227–266 (1974).
29. Hare, D., Austin, C., Doble, P. & Arora, M. Elemental bio-imaging of trace elements in teeth using laser ablation–inductively coupled plasma–mass spectrometry. *J. Dent.* **39**, 397–403 (2011).
30. Agency for Toxic Substances and Disease Registry (ATSDR). *Toxicological Profile for Barium* (US Department of Health and Human Services, Public Health Service, 2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements L. Reynard, N. Tuross and F. Bidlack provided comments on this project. C. Amarasingwardena and N. Lupoli provided expertise in macaque milk analysis. Fossil samples were provided by M. Toussaint, R. Gruen and M.-H. Moncell. The CHAMACOS study is funded by the US Environmental Protection Agency (RD 83171001 and RD 82670901 to B.E.) and the US National Institutes of Environmental Health Sciences (P01 ES009605 to B.E.). Support for macaque data collection was provided by NSF BCS-0921978 (K.H.); milk samples were made possible through the ARMMS program (Archive of Rhesus Macaque Milk Samples). Histological study of the Sladina Neanderthal was funded by the Max Planck Institute for Evolutionary Anthropology. R.J.-B. is supported by Australian Research Council Discovery Grant (DP120101752) and SCU postdoctoral Fellowship grant. P.D. was supported by Australian Research Council Project Grant (LP100200254) that draws collaborative funding from Agilent Technologies and Kennelec Scientific. M.A. is supported by a National Institute of Environmental Health Sciences grant 4R00ES019597-03. C.A. and M.A. are supported by NHMRC grant APP1028372.

Author Contributions C.A., T.M.S. and M.A. designed the study, undertook the elemental and histological analysis, and wrote the manuscript. A.B. and B.E. designed and analysed the human study. K.H. designed the macaque lactation study and collected samples. R.J.-B. analysed the Payre Neanderthal tooth in the Supplementary Information and assessed diagenetic alteration. C.A., D.J.H., D.B. and P.D. undertook elemental imaging of tooth samples. All authors contributed to the interpretation of the results, in addition to editing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A. (marora@hsph.harvard.edu).

METHODS

Human study participants. We used teeth from children enrolled in the Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS) study in Monterey County, California^{27,31}. Pregnant women in the CHAMACOS cohort were recruited before 20 weeks gestation, and data on breastfeeding and infant formulas used were prospectively collected. Interviews were conducted with participants twice during pregnancy (at the end of the first and second trimesters), immediately postpartum, and when children were approximately 6, 12, 24 and 42 months old. Interviews were conducted in person, either at the study office or in a modified recreational vehicle that was used as a mobile office at the participant's home. All questionnaires were administered in English or Spanish by trained bicultural interviewers, with most interviews (94%) conducted in Spanish. Study instruments were developed in English, translated, and validated by Mexican-American immigrant staff members familiar with the language of the community and of southern Mexico from where many participants migrated.

At the second pregnancy interview (mean = 27 weeks gestation), the participant was asked if she intended to breastfeed her child. At each of the postpartum interviews she was asked if she was currently breastfeeding. At the interview when the mother first answered that she was no longer breastfeeding, she was then asked the child's age when she had completely stopped breastfeeding and the reasons for stopping. Additionally, at the 6-month interview, the mother was asked if her child was receiving formula, and if so, at what age formula had been introduced. At the 12-month interview, she was asked at what age formula, solid foods and cow's milk were each introduced. Duration of exclusive breastfeeding was defined as the period between birth and the age when food or liquid other than breast milk or water was first given. All procedures were reviewed by the University of California at Berkeley Committee for the Protection of Human Subjects. Written informed consent was obtained from parents of all participating children and oral assent was obtained from 7 year olds.

From the 7-year assessment onwards, mothers were asked to bring in a tooth the child had shed. We randomly selected deciduous teeth that were free of obvious defects (caries, hypoplasias, fluorosis, cracks, extensive attrition) from 25 children who fell into one of three categories: exclusively breastfed from birth; initially breastfed with formula introduced within 1–2 months of birth; or exclusively formula fed soon after birth. We prepared ~100–150- μm -thick sections in an axial labio-lingual plane following established methods. Developmental times were assigned to marked shifts in Ba/Ca in tooth sections with histological analyses. We photographed the enamel–dentine junction and the neonatal line in enamel and dentine. We overlaid these photomicrographs on our elemental maps to distinguish pre- and postnatal regions (Fig. 1). In teeth of children whose mothers introduced formula within 1–2 months of birth, we noticed clear high Ba/Ca bands in the postnatally formed dentine some distance from the neonatal line. To assign a developmental time to these zones, we used polarized light microscopy to visualize prominent long-period incremental lines and cross-striations (daily growth increments) in enamel, and measured the distance between consecutive cross-striations to determine the average daily enamel secretion rate. Developmental times were then assigned to different points in enamel and dentine along the enamel–dentine junction.

Macaques. Data and samples were obtained from two mother–infant dyads and two additional juveniles at the California National Primate Research Center, UC

Davis, California. All subjects were housed in large, intact social groups in outdoor corrals (0.2 ha). Mothers received a nutritionally complete commercial diet (Outdoor Monkey Lab Diet, PMI Nutrition, Intl) twice daily. Subjects were part of a larger, on-going study on lactation and infant development⁸. Three times during lactation, at infant age 1, 3–4 and 5–6 months, mothers and infants were relocated for milk collection and morphometric measurements as described in detail elsewhere⁸. In the week previous to milk collection, trained technicians conducted four 10-min focal observations between 8:30 and 12:30 and recorded duration of infant suckling behaviour²⁸. All experimental procedures were conducted in accordance with ethical guidelines and with UC Davis Institutional Animal Care and Use Committee approval. Denticitions were collected opportunistically during animal necropsy as part of the CNPRC Biological Specimens Program.

Neanderthal sample. The Scladina Neanderthal upper first maxillary molar was sectioned and temporally mapped for a previous developmental study that established this individual died at approximately 8 years of age¹³.

Ba measurements in teeth using laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS). We used a New Wave Research UP-213 laser ablation system equipped with a Nd:YAG laser emitting a nanosecond laser pulse in the fifth harmonic with a wavelength of 213 nm. The laser was connected to an Agilent Technologies 7500cs ICP-MS by Tygon tubing. Details of our analytical methods have been published previously²⁹. In brief, the laser beam was rastered along the sample surface in a straight line. A laser spot size of 30 μm , laser scan speed of 60 $\mu\text{m s}^{-1}$ and ICP-MS total integration time of 0.50 s produced data points that corresponded to a pixel size³² of approximately 900 μm^2 . Reported element ratios ($\text{Ba/Ca} \times 10^{-4}$ and $\text{Sr/Ca} \times 10^{-3}$) were calculated from concentrations determined using ⁴³Ca, ⁸⁸Sr and ¹³⁸Ba isotopes and NIST 1486 bone meal as a standard. NIST 1486 was not certified for Ba so an average concentration calculated from determinations in two other studies^{33,34} was used. Diagenetic indicators were quantified using NIST 612 glass standard. Each line of ablation produced a single data file in comma separated value (.csv) format. Data were processed using Interactive Spectral Imaging Data Analysis Software (ISIDAS), a custom-built software tool written using Python programming language. ISIDAS reduced all .csv files into a single, exportable visualization toolkit (.vtk) file format. Images were produced by exporting .vtk files into MayaVi2 (Enthought Inc.), an open source data visualization application. Colour scales were applied using the linear blue–red LUT. Image backgrounds were converted to black (absent from the colour intensity scale) to clarify sample boundaries from the substrate. Elemental maps were rotated and black borders added where needed to align rectangular figure panels.

- Eskenazi, B. *et al.* Association of *in utero* organophosphate pesticide exposure and fetal growth and length of gestation in an agricultural population. *Environ. Health Perspect.* **112**, 1116–1124 (2004).
- Lear, J., Hare, D., Adlard, P., Finkelstein, D. & Doble, P. Improving acquisition times of elemental bio-imaging for quadrupole-based LA-ICP-MS. *J. Anal. At. Spectrom.* **27**, 159–164 (2012).
- Porte, N., Mauerhofer, E. & Denschlag, H. O. Test of multielement analysis of bone samples using instrumental neutron activation analysis (INAA) and anti-Compton spectrometry. *J. Radioanal. Nucl. Chem.* **224**, 103–107 (1997).
- Zaichick, V., Zaichick, S., Karandashev, V. & Nosenko, S. The effect of age and gender on Al, B, Ba, Ca, Cu, Fe, K, Li, Mg, Mn, Na, P, S, Sr, V, and Zn contents in rib bone of healthy humans. *Biol. Trace Elem. Res.* **129**, 107–115 (2009).

De novo mutations in histone-modifying genes in congenital heart disease

Samir Zaidi^{1,2*}, Murim Choi^{1,2*}, Hiroko Wakimoto³, Lijiang Ma⁴, Jianming Jiang^{3,5}, John D. Overton^{1,6,7}, Angela Romano-Adesman⁸, Robert D. Bjornson^{7,9}, Roger E. Breitbart¹⁰, Kerry K. Brown³, Nicholas J. Carriero^{7,9}, Yee Him Cheung¹¹, John Deanfield¹², Steve DePalma³, Khalid A. Fakhro^{1,2}, Joseph Glessner¹³, Hakon Hakonarson^{13,14}, Michael J. Italia¹⁵, Jonathan R. Kaltman¹⁶, Juan Kaski¹², Richard Kim¹⁷, Jennie K. Kline¹⁸, Teresa Lee⁴, Jeremy Leipzig¹⁵, Alexander Lopez^{1,6,7}, Shrikant M. Mane^{1,6,7}, Laura E. Mitchell¹⁹, Jane W. Newburger¹⁰, Michael Parfenov³, Itsik Pe'er²⁰, George Porter²¹, Amy E. Roberts¹⁰, Ravi Sachidanandam²², Stephan J. Sanders^{1,23}, Howard S. Seiden²⁴, Mathew W. State^{1,23}, Sailakshmi Subramanian²², Irina R. Tikhonova^{1,6,7}, Wei Wang^{15,25}, Dorothy Warburton^{4,26}, Peter S. White^{14,15}, Ismee A. Williams⁴, Hongyu Zhao^{1,27}, Jonathan G. Seidman³, Martina Brueckner^{1,28}, Wendy K. Chung^{4,29}, Bruce D. Gelb^{22,24,30}, Elizabeth Goldmuntz^{14,31}, Christine E. Seidman^{3,5,32} & Richard P. Lifton^{1,2,6,7,33}

Congenital heart disease (CHD) is the most frequent birth defect, affecting 0.8% of live births¹. Many cases occur sporadically and impair reproductive fitness, suggesting a role for *de novo* mutations. Here we compare the incidence of *de novo* mutations in 362 severe CHD cases and 264 controls by analysing exome sequencing of parent-offspring trios. CHD cases show a significant excess of protein-altering *de novo* mutations in genes expressed in the developing heart, with an odds ratio of 7.5 for damaging (premature termination, frameshift, splice site) mutations. Similar odds ratios are seen across the main classes of severe CHD. We find a marked excess of *de novo* mutations in genes involved in the production, removal or reading of histone 3 lysine 4 (H3K4) methylation, or ubiquitination of H2BK120, which is required for H3K4 methylation^{2–4}. There are also two *de novo* mutations in *SMAD2*, which regulates H3K27 methylation in the embryonic left-right organizer⁵. The combination of both activating (H3K4 methylation) and inactivating (H3K27 methylation) chromatin marks characterizes ‘poised’ promoters and enhancers, which regulate expression of key developmental genes⁶. These findings implicate *de novo* point mutations in several hundreds of genes that collectively contribute to approximately 10% of severe CHD.

From more than 5,000 probands enrolled in the Congenital Heart Disease Genetic Network Study of the National Heart, Lung, and Blood Institute Paediatric Cardiac Genomics Consortium⁷, we selected 362 parent-offspring trios comprising a child (proband) with severe CHD and no first-degree relative with identified structural heart disease. Probands with an established genetic diagnosis were excluded. There were 154 probands with conotruncal defects, 132 with left ventricular obstruction, 70 with heterotaxy and six with other diagnoses (Supplementary Table 1).

Genomic DNA samples from trios underwent exome sequencing⁸ (see Methods). Targeted bases in each sample were sequenced a mean of 107 times by independent reads, with 96.0% read eight or more times. In parallel, 264 trios comprising unaffected siblings of autism cases and their unaffected parents (Supplementary Table 1) were sequenced in the same facility using the same protocol and were analysed as a control group⁹ (Supplementary Table 2 and Supplementary Fig. 1). Family relationships were confirmed from sequence data in all trios.

High-probability *de novo* variants in probands were identified using a Bayesian quality score (QS; see Methods). Sanger sequencing of 181 putative *de novo* mutations across the QS spectrum demonstrated strong correlation of confirmation with QS ($R^2 = 0.89$), with 100% confirmation of 90 calls with QS > 50 (Supplementary Table 3 and Supplementary Fig. 2). Consequently, *de novo* mutation calls with QS ≥ 50 were included in the study; this set is estimated to include 90% of mutations with QS > 0, with ~100% specificity; 90% of these have the maximum QS of 100 (Supplementary Fig. 3). Sensitivity is further diminished by ~5% owing to bases with very low read coverage. We found 0.88 *de novo* mutations per subject in CHD cases and 0.85 in controls. These mutation rates (1.34 and 1.29×10^{-8} per targeted base) are not significantly different ($P = 0.63$, binomial test) and are similar to previous estimates¹⁰. The set of *de novo* mutations is shown in Supplementary Table 4.

CHD cases and controls had very similar maternal and paternal ages, which had a small effect on the mutation rate (Supplementary Fig. 4). We found no significant effect of geographic ancestry on the mutation rate (Supplementary Fig. 5). The number of *de novo* mutations per subject closely approximated the Poisson distribution, providing no evidence for mutation clustering (Supplementary Fig. 6).

¹Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA. ²Howard Hughes Medical Institute, Yale University, Connecticut 06510, USA. ³Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Pediatrics, Columbia University Medical Center, New York, New York 10032, USA. ⁵Howard Hughes Medical Institute, Harvard University, Boston, Massachusetts 02115, USA. ⁶Yale Center for Mendelian Genomics, New Haven, Connecticut 06510, USA. ⁷Yale Center for Genome Analysis, Yale University, New Haven, Connecticut 06511, USA. ⁸Steven and Alexandra Cohen Children's Medical Center of New York, New Hyde Park, New York 11040, USA. ⁹Department of Computer Science, Yale University, New Haven, Connecticut 06511, USA. ¹⁰Department of Cardiology, Children's Hospital Boston, Boston, Massachusetts 02115, USA. ¹¹Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York 10032, USA. ¹²Department of Cardiology, University College London, Great Ormond Street Hospital, London WC1N 3JH, UK. ¹³Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. ¹⁴Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ¹⁵The Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. ¹⁶National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁷Section of Cardiothoracic Surgery, University of Southern California Keck School of Medicine, Los Angeles, California 90089, USA. ¹⁸Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York 10032, USA. ¹⁹Division of Epidemiology, Human Genetics and Environmental Sciences, University of Texas School of Public Health, Houston, Texas 77030, USA. ²⁰Department of Computer Science, Columbia University, New York, New York 10032, USA. ²¹Department of Pediatrics, University of Rochester Medical Center, The School of Medicine and Dentistry, Rochester, New York 14611, USA. ²²Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²³Program on Neurogenetics, Child Study Center, Department of Psychiatry, Yale University, New Haven, Connecticut 06510, USA. ²⁴Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²⁵Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey 07102, USA. ²⁶Department of Pathology, Columbia University Medical Center, New York, New Jersey 10032, USA. ²⁷Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut 06510, USA. ²⁸Department of Pediatrics Yale University School of Medicine, New Haven, Connecticut 06510, USA. ²⁹Department of Medicine, Columbia University Medical Center, New York, New York 10032, USA. ³⁰The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ³¹Division of Cardiology, The Children's Hospital of Philadelphia, The University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. ³²Cardiovascular Division, Brigham & Women's Hospital, Harvard University, Boston, Massachusetts 02115, USA. ³³Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

*These authors contributed equally to this work.

Table 1 | *De novo* mutations in genes with high expression in developing heart in CHD probands and controls

Mutations in genes in top quartile of expression at E14.5	Total no. <i>de novo</i> mutations		<i>De novo</i> mutations/subject		Odds ratio cases: cont (95% CI)†	P value††
	CHD 362 trios	Controls 264 trios	CHD 362 trios	Controls 264 trios		
Silent	21	21	0.06	0.08	NA	0.35
Non-conserved missense	27	17	0.07	0.06	1.59 (0.67–3.74)	0.76
Silent and protein changing	102	53	0.28	0.20	NA	0.05
All protein changing	81	32	0.22	0.12	2.53 (1.22–5.25)	0.003
Conserved missense	39	13	0.11	0.05	3.00 (1.25–7.17)	0.01
Conserved and damaging protein altering	54	15	0.15	0.06	3.6 (1.57–8.28)	0.0005
Damaging	15	2	0.04	0.01	7.50 (1.52–36.95)	0.01

† The odds ratio is the ratio of protein-altering to silent variants in cases divided by the corresponding ratio in controls.

†† P values compare the number of variants in each category between cases and controls using a two-tailed binomial exact test.

CI, confidence interval; NA, not applicable.

Genes contributing to CHD should be expressed in the developing heart/anlagen or tissues that provide developmental cues. We used RNA sequencing of mouse heart at embryonic day (E)14.5 (Methods) to partition 16,676 genes with identified human–mouse orthologues into the top quartile of expression (4,169 genes with high heart expression, HHE; threshold, >40 reads per million mapped reads (r.p.m.)) and the bottom 75% (12,507 with lower heart expression, LHE). The HHE set included regulatory genes known to be expressed at this stage such as *Gata4*, *Nkx2-5* and *Tbx5*.

We found a significant increase in the rate of protein-altering *de novo* mutations in HHE genes in patients with CHD compared to controls ($P = 0.003$, binomial test, odds ratio = 2.53, Table 1). Because it is unlikely that all such *de novo* mutations alter protein function, we enriched for deleterious *de novo* mutations, first removing missense mutations at weakly conserved positions among vertebrate orthologues (two or more species with substitutions, median seven), then removing missense mutations at highly conserved positions (zero or one species with substitution, 72% with zero), leaving only damaging mutations (premature termination, splice site and frameshift). This produced successive increases in the odds ratios to 3.60 and 7.50, with significant differences between cases and controls in each group (Table 1 and Fig. 1a). The rise in odds ratio with increasing stringency was significant ($P = 0.001$, logistic model regression). Other predictors of deleterious mutations, such as PolyPhen-2, yielded similar results (probably deleterious missense mutations plus damaging mutations; $P = 0.0007$, binomial test). Similar results were found when genes were partitioned across a range of expression thresholds in the developing heart (Supplementary Table 5) and also when analyses used heart RNA expression from E9.5 (Supplementary Table 6). By contrast, there was no significant difference in mutation frequency in CHD cases versus controls among LHE genes, with odds ratios near or <1 in all comparisons (Fig. 1a and Supplementary Table 7). Analysis comparing the presence or absence of *de novo* mutations in each case and control yielded similar results (Supplementary Table 8 and Supplementary Fig. 7). Examination of subjects with left ventricular obstruction, conotruncal defects and heterotaxy demonstrated similarly increased odds ratios for each group (Supplementary Table 9).

Comparison of *de novo* mutation frequencies in HHE genes versus LHE genes in the CHD cohort also revealed a significantly greater rate in HHE genes, again with odds ratios increasing with increasingly stringent filters (Fig. 1b and Supplementary Table 7). By contrast, controls showed no significant difference in mutation frequencies in HHE versus LHE, again with all odds ratios near or <1 (Fig. 1b and Supplementary Table 7).

Notably, examination of genes mutated in the CHD set revealed eight involved in the production, removal or reading of methylation of H3K4 (H3K4me). Interestingly, three genes in this pathway (*MLL2*, *KDM6A*, *CHD7*) have previously been implicated in rare syndromic CHD^{11,12}. In Gene Ontology analysis (<http://david.abcc.ncifcrf.gov/>) of the 249 protein-altering *de novo* mutations in CHD probands, the H3K4me pathway was the only gene set with significant enrichment ($P = 4 \times 10^{-7}$, modified Fisher's exact test, $P = 4 \times 10^{-4}$ after

Bonferroni correction; see Methods). The number of mutations in this gene set expected by chance was one and controls showed none.

H3K4me is an activating mark found in promoters/enhancers of key developmental genes⁶. Early in development 'poised' promoters/enhancers have both activating H3K4me marks and inactivating H3K27me marks; these promoters/enhancers and their target genes are selectively activated by modification of these marks in different lineages. Mutations in these genes (Table 2 and Fig. 2) included 27% of the damaging mutations in the HHE gene set. Mutated genes included *MLL2* (frameshift) and *WDR5* (missense), components of the MLL2 H3K4 N-methyltransferase complex²; *KDM5A* (missense) and *KDM5B* (splice donor), both H3K4 demethylases³; and *CHD7* (premature termination), an ATP-dependent helicase that binds H3K4me sites¹². There were also *de novo* mutations in *RNF20* (premature termination) and *UBE2B* (missense), components of a histone H2BK120 ubiquitination complex and in *USP44* (missense), encoding a histone H2B deubiquitinase⁴. Ubiquitination at H2BK120 is required for H3K4 methylation².

Interestingly, *SMAD2* is mutated twice (splice site, conserved missense), a finding unlikely to occur by chance ($P = 0.015$, Monte Carlo simulation) (Table 2). *SMAD2* is asymmetrically phosphorylated downstream of NODAL signalling in the embryonic left–right organizer, resulting in *SMAD2* binding to chromatin, recruitment of JMJD3 and demethylation of H3K27me, enabling transcriptional activation at poised sites⁵. Additional genes of note (Table 2) include *SUV420H1*

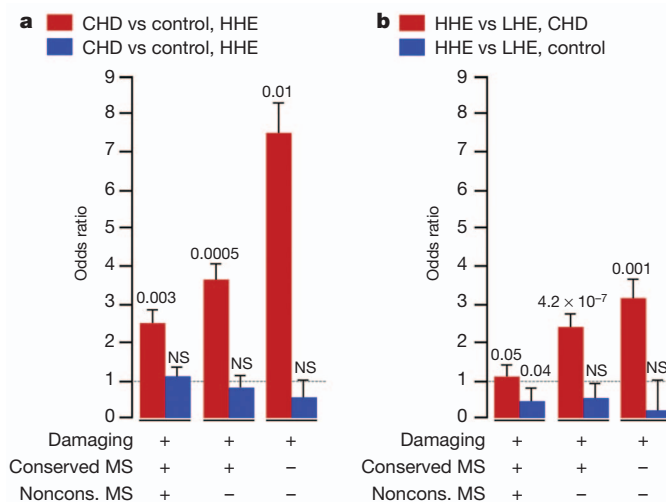


Figure 1 | Enrichment of nonsynonymous *de novo* mutations in heart-expressed genes. **a**, Odds ratios, standard errors and P values (two-tailed binomial exact test) are shown comparing incidence of classes of *de novo* mutations in CHD cases versus controls for genes in top 25% (red bars) and bottom 75% (blue bars) of expression at E14.5 in the developing heart. **b**, Odds ratios for incidence of mutations in genes in top 25% versus bottom 75% of expression in CHD cases (red bars) and controls (blue bars). Damaging denotes premature termination, frameshift or splice site mutations; conserved MS and noncons. MS denote mutations at highly or poorly conserved positions, respectively. NS, not significant.

Table 2 | Genes of interest with *de novo* mutations in probands

ID	Gene	Mutation	Dx	Other structural/neuro/ht-wt
1-00596	<i>MLL2</i> †	p.Ser1722Arg fs*9	LVO	Y/Y/N
1-00853	<i>WDR5</i> †	p.Lys7Gln	CTD	N/Y/N
1-00534	<i>CHD7</i> †	p.Gln1599*	CTD	Y/Y/Y
1-00230	<i>KDM5A</i> †	p.Arg1508Trp	LVO	N/N/Y
1-01965	<i>KDM5B</i> †	p.IVS12 + 1 G>A	LVO	N/N/Y
1-01907	<i>UBE2B</i> †	p.Arg8Thr	CTD	N/N/N
1-00075	<i>RNF20</i> †	p.Gln83*	HTX	Y/Y/Y
1-01260	<i>USP44</i> †	p.Glu71Asp	LVO	N/N/N
1-02020	<i>SMAD2</i> ††	p.IVS6 + 1 G>A	HTX	Y/N/N
1-02621	<i>SMAD2</i> ††	p.Trp244Cys	HTX	Y/NA/N
1-01451	<i>MED20</i>	p.IVS2 + 2 T>C	HTX	N/Y/Y
1-01151	<i>SUV420H1</i>	p.Arg143Cys	CTD	N/Y/N
1-00750	<i>HUWE1</i>	p.Arg3219Cys	LVO	N/Y/N
1-00577	<i>CUL3</i>	p.Iso145Phe fs*23	LVO	Y/Y/N
1-00116	<i>NUB1</i>	p.Asp310His	CTD	Y/Y/Y
1-01828	<i>DAPK3</i>	p.Pro193Leu	CTD	N/N/NA
1-03151	<i>SUPT5H</i>	p.Glu451Asp	LVO	N/NA/N
1-00455	<i>NAA15</i>	p.Lys336Lys fs*6	HTX	Y/Y/N
1-00141	<i>NAA15</i>	p.Ser761*	CTD	N/NA/Y
1-01138	<i>USP34</i>	p.Leu432Pro	LVO	N/NA/N
1-00448	<i>NF1</i>	p.IVS6 + 4 del A	CTD	N/NA/N
1-00802	<i>PTCH1</i>	p.Arg831Gln	LVO	N/NA/N
1-02458	<i>SOS1</i>	p.Thr266Lys	Other	Y/Y/Y
1-02952	<i>PITX2</i>	p.Ala47Val	LVO	N/NA/N
1-01913	<i>RAB10</i>	p.Asn112Ser	Other	N/NA/N
1-00638	<i>FBN2</i>	p.Asp2191Asn	CTD	N/NA/N
1-00197	<i>BCL9</i>	p.Met1395Lys	LVO	N/NA/N
1-02598	<i>LRP2</i>	p.Glu4372Lys	HTX	N/NA/N

Gene symbols are as in NCBI RefSeq database. Other structural/neuro/ht-wt denotes presence (Y) or absence (N) of other structural abnormalities, impaired cognitive speech or motor development, and height (ht) and/or weight (wt) less than 5th percentile for age, respectively. Further clinical details in Supplementary Tables 10 and 11. Associated syndromes: *MLL2*, Kabuki syndrome; *CHD7*, CHARGE syndrome; *CUL3*, pseudohypoparathyroidism, type 2E.

* Premature termination mutation.

† Gene involved in production, removal or reading of H3K4 methylation mark.

†† Gene involved in removal of H3K27 methylation mark.

Del, deletion; Dx, diagnosis; fs, frameshift mutation; fs*n, frameshift mutation followed by premature termination *n* codons later; NA, data not available.

(missense), encoding a histone H4 methylase; *MED20* (splice site), a component of the mediator complex; *HUWE1* (missense), a ubiquitin ligase targeting histones and TP53; *CUL3* (frameshift), a scaffold for assembly of many RING ubiquitin ligases⁸; and *NUB1* (missense), which inhibits NEDD8, a cofactor for cullin-based ubiquitin ligases. Last, *NAA15*, an *N*-acetyltransferase¹³, had two damaging mutations, unlikely a chance event ($P = 0.01$, Monte Carlo simulation). Among the 17 above genes, ten have no damaging variants and seven have one to five among >9,500 exomes in National Heart, Lung, and Blood Exome Sequencing Project, 1000 Genomes and Yale exome databases.

Phenotypes of the eight patients with *de novo* mutations in the H3K4me pathway revealed diverse cardiac phenotypes (Table 2 and Supplementary Table 10). Other structural, neurodevelopmental and growth abnormalities were common. In addition, consistent with a role in left-right axis determination⁵, both patients with *SMAD2* mutations had dextrocardia with unbalanced complete atrioventricular canal and pulmonary stenosis. For other genes mutated more than once (for example, *NAA15*), probands had dissimilar cardiac phenotypes (Supplementary Table 11).

Before initiating exome sequencing, we defined a set of 277 candidate CHD genes (Supplementary Table 12) from human and model system studies. There were 13 CHD probands with *de novo* mutations in these genes (Table 2 and Supplementary Table 13), more than expected by chance ($P = 7 \times 10^{-4}$, Monte Carlo simulation) or in controls ($n = 1$; $P = 0.006$, binomial test). This set included several genes known to cause Mendelian CHD; however, affected subjects lacked cardinal disease manifestations or had atypical cardiac features. For example, the patient with the *CHD7* mutation had none of the main criteria (coloboma, choanal atresia or hypoplastic semicircular canals) for CHARGE syndrome¹². Similarly, the patient with the *MLL2* mutation was not prospectively diagnosed with Kabuki syndrome; however, re-evaluation at age 2 after sequencing identified characteristic facial features. Additionally, a patient with an *NF1* mutation had a

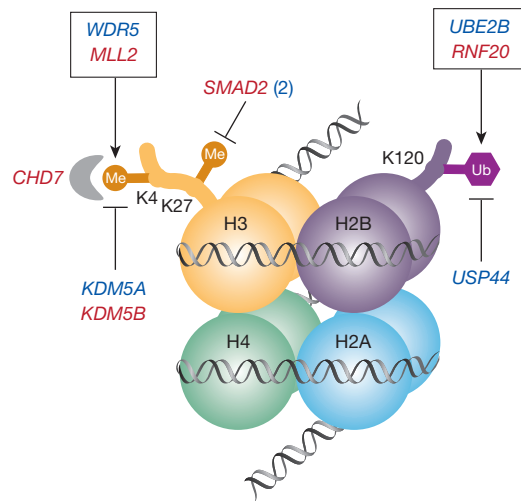


Figure 2 | *de novo* mutations in the H3K4 and H3K27 methylation pathways. Nucleosome with histone octamer and DNA, with H3K4 methylation bound by CHD7, H3K27 methylation and H2BK120 ubiquitination is shown. Genes mutated in CHD that affect the production, removal and reading of these histone modifications are shown; genes with damaging mutations are shown in red, those with missense mutations are shown in blue. *SMAD2* (2) indicates there are two patients with a mutation in this gene. Genes whose products are found together in a complex are enclosed in a box.

complex conotruncal defect, an unusual finding in neurofibromatosis. These findings support variable expressivity and a broader phenotypic spectrum resulting from mutations at known disease loci. Other genes of interest in this set included *RAB10* and *BCL9*, identified as candidates by rare *de novo* copy-number variants¹⁴.

Our results implicate *de novo* point/insertion-deletion (indel) mutations that by chance occur in genes required for normal heart development in the pathogenesis of diverse CHDs. Consistent with this inference, genes with damaging and conserved missense mutations in CHD probands showed higher expression in E14.5 mouse heart compared to controls (Supplementary Fig. 8; median 45 versus 16 r.p.m.; $P = 5 \times 10^{-4}$, Wilcoxon signed-rank test), whereas expression of genes with silent mutations show no significant difference (median 21 versus 19 r.p.m.; $P = 0.7$, Wilcoxon signed-rank test). Expression at E9.5 shows similar results (Supplementary Fig. 8). The increased mutation burden of HHE genes in cases is not due to a higher intrinsic mutation rate of these genes because the rate is significantly higher than in controls; moreover, there is no significant difference in mutation rate between HHE and LHE genes in controls. Further, partitioning genes into analogous high- and low-expression groups for four control adult tissues (brain, heart, liver and lung) showed no significant differences in mutation burden between cases and controls or between high- and low-expression groups (Supplementary Fig. 9).

From the increased fraction of patients with protein-altering mutations in HHE genes in CHD patients (0.22) versus controls (0.12), we estimate that such mutations have a role in about 10% of these patients (95% confidence interval, 5–15%). This could be somewhat underestimated, as mutation detection is incomplete, analysis is limited to genes with identified mouse orthologues, and the HHE set may not include all trait loci. Similarly, the observed odds ratios may be somewhat underestimated as not all mutations in cases are likely to confer risk.

These findings establish that mutations in many genes in the H3K4me-H3K27me pathway disrupt cardiac development and are consistent with previous evidence implicating these chromatin marks in regulating key developmental genes⁶, including those involved in cardiac development^{15,16}. Targeted sequencing in larger CHD cohorts will enable assessment of the role of each individual gene in this pathway. These findings imply dosage sensitivity for these chromatin marks in CHD, similar to recent findings implicating haploinsufficiency for chromatin modifying/remodelling genes in diverse

cancers^{17,18}. Investigation of the consequences of these mutations on specific enhancers/promoters and the genes they regulate will probably provide further insight into the CHD pathogenesis.

The demonstration that point/indel mutations contribute to ~10% of CHD patients and the finding that six genes were mutated twice (Supplementary Table 11) enables an estimate of the size of the gene set that contributes to these CHDs (see Methods). The point-wise estimate is 401 genes (95% confidence interval, 197–813), indicating that many more CHD-related genes and pathways remain to be discovered.

Exome sequencing of probands with autism have revealed broadly similar results: *de novo* mutations in a large set of genes occur in a significant fraction of patients, with relatively high odds ratios for damaging mutations in genes expressed in the brain^{9,19–21}. Most interestingly, CHD8, which like CHD7 reads H3K4me marks, is frequently mutated in autism²², raising the question of whether the H3K4me pathway may have a role in many congenital diseases. Among 249 protein-altering *de novo* mutations in CHD (Supplementary Table 4) and 570 such mutations in autism^{9,19,20,23}, there were two genes, *CUL3* and *NCKAP1*, with damaging mutations in both CHD and autism and none in controls ($P = 0.001$, Monte Carlo simulation), and several others with mutations in both (for example, *SUV40H1* and *CHD7*). Similarly, rare copy-number variants at 22q11.2, 1q21 and 16p11 are found in patients with autism, CHD or both diseases^{24–26}. These observations suggest variable expressivity of mutations in key developmental genes. Identification of the complete set of these developmental genes and the full spectrum of the resulting phenotypes will likely be important for patient care and genetic counselling.

Our findings do not resolve the pathogenesis of most CHD cases. Rare and *de novo* copy-number variants seem to account for a small fraction^{14,27}; rare or common transmitted variants are also expected to make significant contributions. Additionally, considering the role of H3K4me and H3K27me marks in promoter/enhancer regulation, non-coding mutations cannot be dismissed. Last, evidence of dosage sensitivity of many chromatin-modifying genes raises the possibility that environmental perturbations of these pathways in critical developmental windows might phenocopy the effects of these mutations.

METHODS SUMMARY

De novo mutations in a cohort of 362 probands with CHD and 264 unaffected subjects were identified by exome sequencing of parent–offspring trios. Gene expression in mouse heart at E14.5 was quantitated by RNA sequencing, and genes in the top quartile of expression were identified. The frequency of *de novo* mutations in genes with higher expression in developing heart was compared in CHD cases and controls. Enrichment of mutations in particular pathways was examined using Gene Ontology.

Full Methods and any associated references are available in the online version of the paper.

Received 16 January; accepted 2 April 2013.

Published online 12 May 2013.

1. Reller, M. D., Strickland, M. J., Riehle-Colarusso, T., Mahle, W. T. & Correa, A. Prevalence of congenital heart defects in metropolitan Atlanta, 1998–2005. *J. Pediatr.* **153**, 807–813 (2008).
2. Shilatfard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).
3. Pedersen, M. T. & Helin, K. Histone demethylases in development and disease. *Trends Cell Biol.* **20**, 662–671 (2010).
4. Fuchs, G. et al. RNF20 and USP44 regulate stem cell differentiation by modulating H2B monoubiquitylation. *Mol. Cell* **46**, 662–673 (2012).
5. Dahle, Ø., Kumar, A. & Kuehn, M. R. Nodal signaling recruits the histone demethylase Jmjd3 to counteract polycomb-mediated repression at target genes. *Sci. Signal.* **3**, ra48 (2010).
6. Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
7. Pediatric Cardiac Genomics Consortium. The Congenital Heart Disease Network Study (CHD GENES): rationale, design and early results. *Circ. Res.* **112**, 698–706 (2013).
8. Boyden, L. M. et al. Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. *Nature* **482**, 98–102 (2012).

9. Sanders, S. J. et al. *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
10. Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nature Rev. Genet.* **13**, 745–753 (2012).
11. Lederer, D. et al. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. *Am. J. Hum. Genet.* **90**, 119–124 (2012).
12. Vissers, L. E. et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nature Genet.* **36**, 955–957 (2004).
13. Gendron, R. L., Adams, L. C. & Paradis, H. Tubedown-1, a novel acetyltransferase associated with blood vessel development. *Dev. Dyn.* **218**, 300–315 (2000).
14. Greenway, S. C. et al. *De novo* copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nature Genet.* **41**, 931–935 (2009).
15. Wamstad, J. A. et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220 (2012).
16. Paige, S. L. et al. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151**, 221–232 (2012).
17. Ceol, C. J. et al. The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* **471**, 513–517 (2011).
18. Sausen, M. et al. Integrated genomic analyses identify *ARID1A* and *ARID1B* alterations in the childhood cancer neuroblastoma. *Nature Genet.* **45**, 12–17 (2013).
19. O’Roak, B. J. et al. Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
20. Iossifov, I. et al. *De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
21. Neale, B. M. et al. Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
22. O’Roak, B. J. et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
23. Kong, A. et al. Rate of *de novo* mutations and the importance of father’s age to disease risk. *Nature* **488**, 471–475 (2012).
24. Vorstman, J. A., Breetvelt, E. J., Thode, K. I., Chow, E. W. & Bassett, A. S. Expression of autism spectrum and schizophrenia in patients with a 22q11.2 deletion. *Schizophr. Res.* **143**, 55–59 (2013).
25. Mefford, H. C. et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* **359**, 1685–1699 (2008).
26. Ghebranious, N., Giampietro, P. F., Wesbrook, F. P. & Rezakalla, S. H. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am. J. Med. Genet.* **143A**, 1462–1471 (2007).
27. Soemedi, R. et al. Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am. J. Hum. Genet.* **91**, 489–501 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors are grateful to the patients and families who participated in this research. We thank the following team members for contributions to patient recruitment: D. Awad, K. Celia, D. Etwaru, R. Korsin, A. Lanz, E. Marquez, J. K. Sond, A. Wilpers, R. Yee (Columbia Medical School); K. Boardman, J. Geva, J. Gorham, B. McDonough, A. Monafio, J. Stryker (Harvard Medical School); N. Cross (Yale School of Medicine); S. M. Edman, J. L. Garbarini, J. E. Tusi, S. H. Woyciechowski (Children’s Hospital of Philadelphia); J. Ellashek and N. Tran (Children’s Hospital of Los Angeles); K. Flack (University College London); D. Gruber, N. Stellato (Steve and Alexandra Cohen Children’s Medical Center of New York); D. Guevara, A. Julian, M. Mac Neal, C. Mintz (Icahn School of Medicine at Mount Sinai); and E. Taillie (University of Rochester School of Medicine and Dentistry). We also thank V. Spottow, P. Andrea, K. Pavlik and M. Sotiropoulos for their expert production of exome sequences. We thank B. Bernstein and R. Ryan (Massachusetts General Hospital) and B. Bruneau (Gladstone Institute and University of California, San Francisco) for discussions. This work was supported by the National Institutes of Health (NIH) National Heart, Lung, and Blood Institute (NHLBI) Pediatric Cardiac Genomics Consortium (U01-HL098188, U01-HL098147, U01-HL098153, U01-HL098163, U01-HL098123, U01-HL098162) and in part by the Simons Foundation for Autism Research and the NIH Centers for Mendelian Genomics (5U54HG006504).

Author Contributions Study design: M.B., W.K.C., B.D.G., E.G., H.H., J.R.K., R.P.L., L.E.M., J.G.S., C.E.S., D.W., P.S.W.; cohort ascertainment, phenotypic characterization and recruitment: R.E.B., M.B., W.K.C., J.D., B.D.G., E.G., J.K., R.K., T.L., J.W.N., G.P., A.R.-A., H.S.S., C.E.S., I.A.W.; informatics/data management: R.D.B., R.E.B., N.J.C., M.C., S.D., J.G., H.H., M.J.L., J.L., A.L., S.M.M., J.D.O., M.P., A.E.R., J.G.S., W.W., P.S.W., S.Z.; exome sequencing production: J.D.O., A.L., R.P.L., S.M.M., M.W.S., I.R.T.; *de novo* mutation validation: W.K.C., L.M.; exome sequencing analysis: K.K.B., Y.H.C., M.C., S.D., K.A.F., J.G., J.K.K., R.P.L., I.P., R.S., S.J.S., J.G.S., C.E.S., S.S., W.W., S.Z.; RNA sequence production/analysis: J.J., M.P., C.E.S., J.G.S., H.W.; statistical analysis: M.C., R.P.L., I.P., A.E.R., C.E.S., J.G.S., S.Z., H.Z.; writing of manuscript: M.B., M.C., W.K.C., B.D.G., E.G., J.R.K., R.P.L., C.E.S., S.Z. Co-senior authors: M.B., W.K.C., B.D.G., E.G., C.E.S. and R.P.L.

Author Information Messenger RNA and protein sequences are available in the RefSeq database (<http://www.ncbi.nlm.nih.gov/refseq/>) under accession numbers listed in Supplementary Table 4; mutation data are available at dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>) under batch accession 1059065. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NHLBI. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. (martina.brueckner@yale.edu), W.K.C. (wkc15@cumc.columbia.edu), B.D.G. (bruce.gelb@mssm.edu), E.G. (goldmuntz@email.chop.edu), C.E.S. (csediman@genetics.med.harvard.edu) or R.P.L. (richard.lifton@yale.edu).

METHODS

Patient cohorts. Probands with or without parents were recruited from nine centres in the United States and the United Kingdom into the Congenital Heart Disease Genetic Network Study of the Paediatric Cardiac Genomics Consortium (CHD genes: ClinicalTrials.gov identifier NCT01196182)⁷. The protocol was approved by the Institutional Review Boards of Boston Children's Hospital, Brigham and Women's Hospital, Great Ormond Street Hospital, Children's Hospital of Los Angeles, Children's Hospital of Philadelphia, Columbia University Medical Center, Icahn School of Medicine at Mount Sinai, Rochester School of Medicine and Dentistry, Steven and Alexandra Cohen Children's Medical Center of New York, and Yale School of Medicine. Written informed consent was obtained from each participating subject or their parent/guardian. Probands were selected for severe CHD (excluding isolated ventricular septal defects, atrial septal defects, patent ductus arteriosus or pulmonic stenosis), availability of both parents and absence of any CHD in first-degree relatives. Cardiac diagnoses were obtained from review of echocardiogram, catheterization and operative reports; extracardiac findings were extracted from medical records. Controls were from 264 previously studied quartets that included one offspring with autism, an unaffected sibling and unaffected parents, all recruited with written informed consent by the Simons Foundation Autism Research Initiative²⁸. Parents and their unaffected sibling from this cohort were analysed in the current study.

Exome sequencing. Trios were sequenced at the Yale Center for Genome Analysis following the same protocol. Genomic DNA from venous blood was captured with the NimbleGen v2.0 exome capture reagent (Roche) and sequenced (Illumina HiSeq 2000, 75 base-paired end reads). Reads were mapped to the reference genome using ELANDv2. Single-nucleotide variants and indel calls were assigned a QS using SAMtools⁸ and annotated for novelty using dbSNP, build 135, 1000 Genomes (May 2011 release) and the Yale Exome Database, for impact on encoded proteins and conservation of variant position.

Identification and confirmation of *de novo* mutations. Heterozygous single nucleotide variants and indels in the proband that showed SAMtools QS ≥ 60 and 600, respectively, and rare non-reference calls in both parents were selected. Read plots of all putative indels were visually inspected in trio members to eliminate false calls. A Bayesian algorithm was used to assist *de novo* mutation calls. Elements included probability of the proband being heterozygous at the test position; probability that parents are homozygous for the reference allele, given frequency of reference and non-reference reads and probability of heterozygosity in offspring; probability that a variant is *de novo* given its population frequency. Resulting Bayesian QSs were scaled from 0 to 100. Their correlation with bona fide *de novo* mutations was determined by Sanger sequencing of PCR amplicons harbouring 181 putative mutations distributed across the Bayesian QS spectrum. Additionally, all six *de novo* indels with Bayesian QS > 50 in the HHE gene set were tested and confirmed by Sanger sequencing.

RNA sequencing and analysis. Hearts from E14.5 mouse embryos (strain 129/SvEv) were isolated, rinsed and immersed in RNALater. Left and right atria, left ventricle (with interventricular septum, aortic and mitral valves) and right ventricle (with pulmonary and tricuspid valves) were dissected. Chamber-specific RNAs were extracted and pooled from five embryos, selected with oligo-dT, copied into double-stranded DNA and ligated to adaptors. 150–250 base-pair fragments were isolated after acrylamide gel electrophoresis, amplified and sequenced (Illumina HiSeq 2000), with >40 million paired-end 50-base reads per library as previously described²⁹. Reads were aligned to the mouse genome (mm9)³⁰ and r.p.m. was determined. The average r.p.m. of each gene from each chamber was used as the measure of heart expression. RNA from atria, ventricle and truncus/outflow tract at E9.5 was prepared, sequenced and analysed by an analogous approach. RNA sequencing of control human adult tissues—lung, liver, heart and brain—from the Illumina Human Body Map (<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/?query=illumina+human+body+map>) was similarly performed and analysed as r.p.m. per kilobase of transcript.

Principal component analysis. The EIGENSTRAT program was used to compare single-nucleotide polymorphisms (SNPs) genotypes of probands and individuals of known ancestry in HapMap3 (<http://hapmap.ncbi.nlm.nih.gov/>). SNPs with minor allele frequency (MAF) $>5\%$ without significant linkage disequilibrium with other SNPs were analysed. The results of analysis correctly distinguished ancestry groups in HapMap3 samples; ancestries of CHD subjects were assigned accordingly.

Statistical analyses. The significance of mutation frequency differences between groups was tested with two-tailed binomial exact tests; two-tailed Fisher's exact tests assessed differences in numbers of patients with one or more *de novo* mutations; tests among three groups was by Chi-square analysis. Gene expression at E14.5 of genes mutated in cases and controls was compared by Wilcoxon signed-rank test. Correlation of mutation rate and parental age was tested by Pearson's correlation. The expected number of genes with more than one *de novo* mutation was determined by Monte Carlo simulation (10^8 iterations) specifying the total number of protein-altering mutations and 21,000 genes of observed coding length. Analogous approaches were used to determine probabilities of any gene having ≥ 2 damaging mutations, ≥ 1 damaging and ≥ 1 mutation at a conserved position, and ≥ 13 genes mutated in both CHD and autism. The fit to the Poisson distribution of the observed numbers of *de novo* mutations per subject was assessed by Chi-square test.

Overrepresentation of *de novo* mutations in the H3K4me pathway and the presence of significant enrichment of other gene pathways was tested by Gene Ontology analysis, using a modified Fisher's exact test with Bonferroni correction as implemented in DAVID (<http://david.abcc.ncifcrf.gov/>). Input was all genes with protein-altering *de novo* mutations in CHD or control subjects, and all genes sequenced. The H3K4me gene set was: *CHD8*, *MLL3*, *SETD7*, *WHSC1L1*, *CDC73*, *WHSC1*, *SETD1A*, *MLL2*, *KDM5A*, *MLL4*, *MLL5*, *UBE2B*, *ASH1L*, *SETD1B*, *MLL*, *LEO1*, *PAF1*, *KDM5C*, *CTR9*, *PRDM9*, *MEN1*, *CHD7*, *RNF20*, *KDM1A*, *RNF40*, *SMYD3*, *KDM6A*, *KDM5B*, *USP44* and *WDR5*. The expected number of mutations in the H3K4me set was calculated from the fraction of the exome-coding region attributable to this gene set and the total number of *de novo* mutations.

Estimating number of genes in which *de novo* mutations contribute to CHD. We addressed this question using the 'unseen species problem'⁹. We infer that the number of probands with nonsynonymous mutations in the HHE set (81) minus the expected number (44; calculated from the number observed in controls) represents the number of subjects in whom *de novo* mutations confer CHD risk (37; 10.0% of probands). The number of genes with >1 protein-altering *de novo* mutation (six) minus the most likely number expected by chance (three) represents risk-associated genes with more than one mutation (three). The number of risk-associated genes (*C*) is estimated as follows:

$$C = c/u + g^2 \times d \times (1-u)/u$$

Where *c* = number of observed risk-associated genes (34), *c*₁ = number of genes mutated once (31), *d* = total number of risk-associated mutations (37), *g* = variation in effect size of individual *de novo* mutations (assumed to be 1, which minimizes underestimation of set size), *u* = $1 - c_1/d$ (probability that newly added mutation hits a previously mutated gene).

$$C = 401.$$

From 95% confidence intervals of the number of risk-associated events, the 95% confidence interval for number of risk genes is calculated as 197–837.

28. Fischbach, G. D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).
29. Christodoulou, D. C., Gorham, J. M., Herman, D. S. & Seidman, J. G. Construction of normalized RNA-seq libraries for next-generation sequencing using the crab duplex-specific nuclease. *Curr. Protoc. Mol. Biol.* 94:4.12.1–4.12.11 (2011).
30. Herman, D. S. *et al.* Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* **366**, 619–628 (2012).

RIP1-driven autoinflammation targets IL-1 α independently of inflammasomes and RIP3

John R. Lukens¹, Peter Vogel², Gordon R. Johnson¹, Michelle A. Kelliher³, Yoichiro Iwakura⁴, Mohamed Lamkanfi^{5,6*} & Thirumala-Devi Kanneganti^{1*}

The protein-tyrosine phosphatase SHP-1 has critical roles in immune signalling, but how mutations in SHP-1 cause inflammatory disease in humans remains poorly defined¹. Mice homozygous for the Tyr208Asn amino acid substitution in the carboxy terminus of SHP-1 (referred to as *Ptpn6*^{spin} mice) spontaneously develop a severe inflammatory syndrome that resembles neutrophilic dermatosis in humans and is characterized by persistent footpad swelling and suppurative inflammation^{2,3}. Here we report that receptor-interacting protein 1 (RIP1)-regulated interleukin (IL)-1 α production by haematopoietic cells critically mediates chronic inflammatory disease in *Ptpn6*^{spin} mice, whereas inflammasome signalling and IL-1 β -mediated events are dispensable. IL-1 α was also crucial for exacerbated inflammatory responses and unremitting tissue damage upon footpad microabrasion of *Ptpn6*^{spin} mice. Notably, pharmacological and genetic blockade of the kinase RIP1 protected against wound-induced inflammation and tissue damage in *Ptpn6*^{spin} mice, whereas RIP3 deletion failed to do so. Moreover, RIP1-mediated inflammatory cytokine production was attenuated by NF- κ B and ERK inhibition. Together, our results indicate that wound-induced tissue damage and chronic inflammation in *Ptpn6*^{spin} mice are critically dependent on RIP1-mediated IL-1 α production, whereas inflammasome signalling and RIP3-mediated necroptosis are dispensable. Thus, we have unravelled a novel inflammatory circuit in which RIP1-mediated IL-1 α secretion in response to deregulated SHP-1 activity triggers an inflammatory destructive disease that proceeds independently of inflammasomes and programmed necrosis.

Mutations in the non-receptor protein tyrosine phosphatase Src homology region 2 (SH2) domain-containing phosphatase-1 (SHP-1) are associated with a spectrum of inflammatory and autoimmune diseases in humans^{4,5}. Similarly, motheaten null and hypomorphic alleles of *Ptpn6*, the gene encoding SHP-1, cause a myeloproliferative disease in mice that is characterized by chronic inflammation^{6–9}. Despite being one of the first *in vivo* genetic models of inflammatory disease, the prevailing mechanism responsible for SHP-1-driven inflammation remains to be formally elucidated. Complete characterization of the molecular mechanism responsible for SHP-1-mediated disease has been hindered by the fact that motheaten mice are immunodeficient, develop devastating pneumonitis and glomerulonephritis, and die by 2–9 weeks of age¹⁰. Recently, a new SHP-1 mutant mouse line (referred to as *Ptpn6*^{spin} mice) that harbours an Tyr208Asn amino acid substitution in the C-terminal SH2 domain of SHP-1 was described³. Mice homozygous for the hypomorphic spin allele develop a chronic inflammatory and autoimmune disease at 8–16 weeks of age that presents as persistent swelling and suppurative inflammation of the cutaneous footpad tissue (Fig. 1a)³. The popliteal lymph nodes that drain the inflamed feet display massive lymphomegaly and are composed of enhanced numbers of both lymphocytes and myeloid cells

(Fig. 1b). In contrast, lymph nodes that drain non-inflamed areas in *Ptpn6*^{spin} mice do not display lymphomegaly (Supplementary Fig. 1). Diseased *Ptpn6*^{spin} mice show elevated levels of circulating cytokines and chemokines that are associated with granulopoiesis and neutrophil recruitment (Fig. 1c). Consistent with augmented production of granulopoietic factors, the inflammatory lesions in the footpads of mutant mice are dominated by neutrophils (Fig. 1d), and neutrophilia ensues in the periphery (Fig. 1e and Supplementary Fig. 2). Moreover, spontaneous disease progression in *Ptpn6*^{spin} mice is characterized by

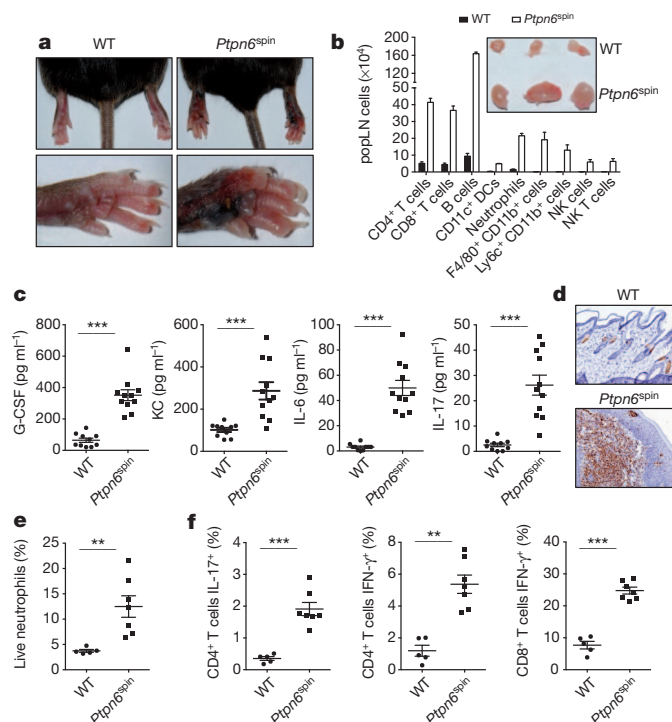


Figure 1 | *Ptpn6*^{spin} mice develop spontaneous footpad inflammation. **a**, Spontaneous induction of footpad swelling (**a**) and lymphomegaly (**b**) in the popliteal lymph nodes (popLN) of *Ptpn6*^{spin} mice at 10–16 weeks of age. **b–f**, Wild-type (WT) and diseased *Ptpn6*^{spin} mice were harvested at 10–12 weeks of age. **b**, Numbers (mean \pm s.e.m.) of popliteal lymph node cells. Inset shows representative pictures of popliteal lymph nodes (original magnification, $\times 1$). **c**, Serum levels of cytokines and chemokines. **d**, Immunohistochemistry staining of neutrophils in the footpads (original magnification, $\times 20$). **e**, Frequency of splenocytes that are neutrophils. **f**, Production of IL-17 and IFN- γ by popliteal lymph node CD4⁺ and CD8⁺ T cells following *in vitro* re-stimulation. Each point represents an individual mouse, and the line represents the mean \pm s.e.m. ** $P < 0.01$, *** $P < 0.001$. DCs, dendritic cells; NK, natural killer.

¹Department of Immunology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ²Animal Resources Center and the Veterinary Pathology Core, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ³Department of Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ⁴Center for Experimental Medicine, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. ⁵Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium. ⁶Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium.

enhanced frequencies of inflammatory T cells that produce high levels of IL-17 and IFN- γ (Fig. 1f and Supplementary Fig. 3) and the accumulation of T cells that exhibit an effector/memory phenotype (CD44^{hi}CD62L^{lo}) (Supplementary Fig. 4). Analysis of mice before the onset of overt disease (4–8 weeks of age) reveals that *Ptpn6*^{spin} mice possess normal numbers of lymphoid and myeloid cells (Supplementary Fig. 5), and do not display perturbations in T-cell development, regulatory T-cell numbers, or T-cell activation status before disease progression (Supplementary Fig. 6). Furthermore, the *Ptpn6*^{spin} mutation does not affect inflammatory cytokine production by peripheral T cells and other immune cells in young mutant mice (Supplementary Fig. 7).

Previous work established that IL-1 receptor (IL-1R) signalling is required for *Ptpn6*^{spin}-mediated inflammatory disease³. However, the molecular mechanisms operating upstream of IL-1R engagement that are responsible for spontaneous induction of inflammatory disease are not known. Inflammasome-driven activation of caspase 1 is increasingly recognized as a central instigator of inflammation and disease pathology through its critical role in the production of bioactive IL-1 β ¹¹. In this context, the NLRP3 inflammasome responds to a multitude of damage-associated danger signals that are associated with autoinflammation¹². To test whether aberrant inflammasome activation is responsible for inducing inflammatory disease in response to defective SHP-1 signalling, *Ptpn6*^{spin} mice were bred to animals that are deficient in the key inflammasome proteins NLRP3 and caspase 1. However, homozygous disruption of neither NLRP3 nor caspase 1 rescued *Ptpn6*^{spin} mice from footpad inflammation (Fig. 2a and Supplementary Fig. 8) and neutrophil infiltration (Fig. 2b). In full agreement, homozygous deletion of the gene encoding IL-1 β also failed to prevent footpad inflammation and granulocyte recruitment in *Ptpn6*^{spin} mice (Fig. 2a, b and Supplementary Fig. 8), nor was excessive inflammatory cytokine production and cutaneous inflammatory disease rescued by genetic deletion of *Tlr4* (Supplementary Fig. 9). In marked contrast, genetic ablation of *Il1a* provided significant

protection from the development of footpad inflammatory disease in *Ptpn6* mutant mice (Fig. 2a, b and Supplementary Fig. 8), which was associated with a return to normal neutrophil numbers (Fig. 2b–d), and reduced generation of IL-17 producing CD4⁺ helper T (T_H17) cells (Fig. 2e). These findings demonstrate that IL-1 α has a central role in SHP-1-mediated disease progression, which proceeds independently of inflammasome activation and IL-1 β secretion.

Given that IL-1 α acts as an alarmin that orchestrates wound-healing responses^{13–15}, we next tested whether defective wound healing might contribute to disease pathogenesis in *Ptpn6*^{spin} mice. To this end, mice were subjected to microabrasion injury on the plantar surfaces of the hind feet, and monitored for incidence of inflammatory responses. Microabrasion-provoked tissue damage induced similar erythema and oedema in wild-type and *Ptpn6*^{spin} mice during the first 48 h. However, inflammation at the wound site was fully resolved in wild-type mice by day 14, whereas *Ptpn6*^{spin} mice developed exacerbated inflammation that was characterized by intense redness and swelling of the affected area (Fig. 3a, b and Supplementary Fig. 10). At day 21, the inability of *Ptpn6*^{spin} mice to curtail wound inflammation ultimately resulted in the development of a persistent and aggravated state of footpad inflammation characterized by severe pustular dermatosis and oedema (Fig. 3a, b and Supplementary Fig. 10). Notably, genetic ablation of IL-1 α production in *Ptpn6*^{spin} mice provided full protection from microabrasion-induced footpad inflammation (Fig. 3a, b and Supplementary Fig. 10). The microabrasion procedure triggered a rapid (4–5 h after wound induction) and potent production of inflammatory cytokines and chemokines in wild-type mice that was further

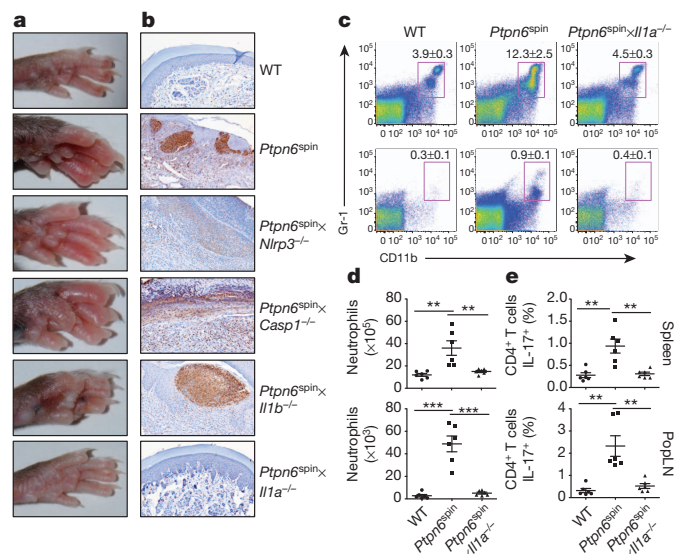


Figure 2 | Deletion of IL-1 α limits *Ptpn6*^{spin}-mediated disease. a, b, Footpad images (a) and neutrophil immunohistochemistry staining (b) of wild-type and *Ptpn6*^{spin} mice that were crossed with mice that are deficient in either NLRP3, caspase 1, IL-1 β , or IL-1 α (original magnification, $\times 20$). c–e, Spleen and popliteal lymph nodes from 12–16-week-old wild-type, *Ptpn6*^{spin} and *Ptpn6*^{spin} × *Il1a*^{-/-} mice. c, Frequencies of neutrophils in the spleen (top panel) and popliteal lymph nodes (bottom panel). Numbers in the FACs plots denote the mean frequencies \pm s.e.m. of cells that are neutrophils. d, Total numbers of neutrophils in the spleen (top panel) and popliteal lymph nodes (bottom panel). e, Production of IL-17 by CD4⁺ T cells following *in vitro* re-stimulation. Data show mean \pm s.e.m. Each point represents an individual mouse, and the line represents the mean \pm s.e.m. ** $P < 0.01$, *** $P < 0.001$.

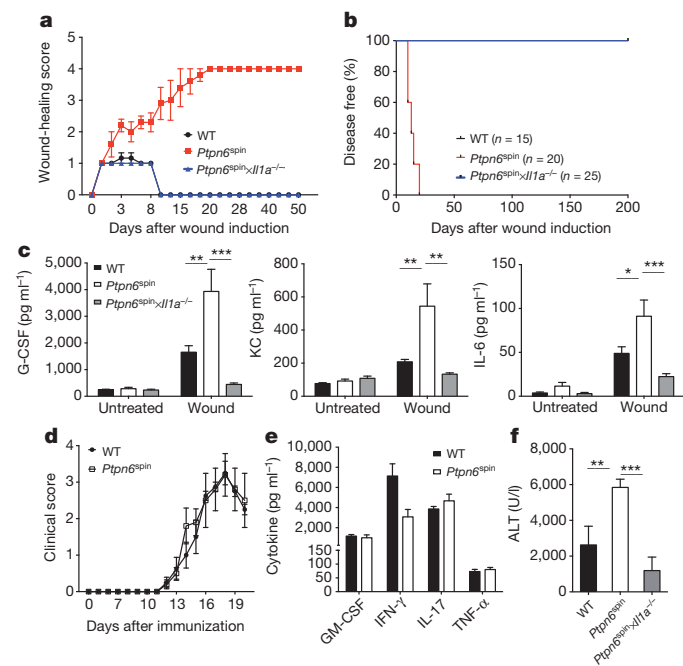


Figure 3 | Exacerbated wound-healing responses contribute to disease in *Ptpn6*^{spin} mice. a–c, Microabrasion injuries were induced on the plantar surfaces of the footpads of wild-type, *Ptpn6*^{spin} and *Ptpn6*^{spin} × *Il1a*^{-/-} mice. a, Clinical scores based on erythema and oedema as described in detail in the Methods section were recorded daily. b, Percentage of disease-free mice over time. c, Serum levels of granulopoiesis-associated factors 5 h after microabrasion stimulation. d, e, Wild-type ($n = 4$) and disease-free PTPN6 mutant mice (5–7 weeks old) ($n = 5$) were immunized with MOG/CFA and pertussis toxin. d, Mean clinical paralysis scores. e, Splenocytes were collected on day 20 and re-stimulated with MOG peptide for 48 h to measure cytokine secretion. f, Wild-type ($n = 7$), disease-free (4–7 weeks of age) *Ptpn6*^{spin} ($n = 9$) and *Ptpn6*^{spin} × *Il1a*^{-/-} ($n = 5$) mice received 250 mg kg⁻¹ of acetaminophen by intraperitoneal injection. The levels of serum alanine aminotransferase (sALT) were measured 18–20 h later by ELISA. All bar graphs show mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

exacerbated in *Ptpn6^{spn}* mice (Fig. 3c). Notably, the enhanced secretion of neutrophilic factors in *Ptpn6^{spn}* mice was fully rescued in *Ptpn6^{spn}* mice lacking IL-1 α (Fig. 3c). Augmented wound-healing responses in *Ptpn6^{spn}* mice were not the result of global aberrations in inflammation because young *Ptpn6^{spn}* mice did not display abnormalities in immune cell composition or inflammatory cytokine production (Supplementary Figs 5–7). Furthermore, the *Ptpn6^{spn}* mutation did not affect the generation of MOG-specific T cells or neuroinflammation during experimental autoimmune encephalomyelitis (EAE) in young mice (Fig. 3d, e and Supplementary Fig. 11). Commensal bacteria are increasingly recognized for their role in the pathogenesis of autoimmune diseases¹⁶, and defects in innate immune signalling were recently shown to alter the intestinal microbiome¹⁷. Because inflammatory disease was suppressed when *Ptpn6^{spn}* mice were derived under germ-free conditions³, we explored the possibility of footpad-associated dysbiosis in *Ptpn6^{spn}* mice. However, total bacterial counts and composition of the footpad-associated microbiome in microabrasion-induced inflammatory skin lesions were comparable in separately housed wild-type and *Ptpn6^{spn}* mice, respectively (Supplementary Fig. 12). Furthermore, we failed to observe enhanced microabrasion-induced granulopoietic cytokine production in wild-type mice that were co-housed with *Ptpn6^{spn}* mice (Supplementary Fig. 13), indicating that the *Ptpn6^{spn}* mutation alters immune responses to normal commensal bacteria rather than modifying the bacterial ecology of inflammatory skin lesions. Notably, *Ptpn6^{spn}* mice also were hypersensitive in the acetaminophen (APAP)-induced liver injury model that is considered a model for sterile autoinflammation and wound-healing responses¹⁸. IL-1 α deletion provided significant protection from APAP-induced liver injury as evidenced by markedly reduced serum alanine aminotransferase levels in APAP-challenged *Ptpn6^{spn} × Il1a^{-/-}* mice (Fig. 3f). Together, these results indicate a critical role for IL-1 α in both sterile and commensal-associated inflammatory and wound-healing responses of *Ptpn6^{spn}* mice.

To determine whether SHP-1 regulates inflammatory responses in haematopoietic or radioresistant cells, bone marrow chimaera mice were generated. Expression of the hypomorphic *Ptpn6^{spn}* allele in the haematopoietic compartment alone promoted spontaneous footpad inflammation (Fig. 4a) concomitant with augmented cytokine production (Fig. 4b) and neutrophilia (Supplementary Fig. 14). In contrast, chimaeric mice bearing the *Ptpn6^{spn}* mutation only in radioresistant cells failed to develop footpad inflammation (data not shown), suggesting that SHP-1 expression in bone-marrow-derived immune cells rather than in non-haematopoietic cells (such as keratinocytes) is critical for induction of the autoinflammatory syndrome. Collectively, these findings suggest that unwarranted IL-1 α release in response to dysregulated SHP-1 activity in haematopoietic cells has a pivotal role in the induction of inflammatory disease. To identify the bone-marrow-derived cell populations that are responsible for *Ptpn6^{spn}*-induced inflammation, we investigated inflammatory responses in isolated macrophages and neutrophils as these cell types have been shown to centrally regulate inflammatory and wound-healing responses¹⁹. The *Ptpn6^{spn}* mutation did not influence inflammatory cytokine production in macrophages (Supplementary Fig. 15). Although *Ptpn6^{spn}* neutrophils produced slightly higher levels of the proinflammatory cytokines granulocyte colony-stimulating factor (G-CSF) and tumour-necrosis factor- α (TNF- α) in response to lipopolysaccharide (LPS) stimulation, production of other pro-inflammatory mediators (KC, IL-6 and IL-1 β) was normal in these cells (Supplementary Fig. 16). We therefore concluded that modest differences in neutrophil-associated cytokine production may contribute to, but are unlikely to account fully for, the marked inflammatory phenotype observed *in vivo*.

The kinase RIP1 is emerging as a key regulator of inflammatory cytokine production and cellular stress^{20,21}. To address the *in vivo* role of RIP1 in exacerbated inflammatory cytokine production, *Ptpn6^{spn}* mice were pre-treated with either vehicle control (PBS), the RIP1 kinase inhibitor necrostatin 1 (Nec1) or the structurally related inactive

Nec1 analogue (iNec)²² before being subjected to microabrasion injury. Unlike iNec, Nec1-mediated *in vivo* inhibition of RIP1 kinase activity markedly attenuated secretion of inflammatory mediators in *Ptpn6^{spn}* mice to levels comparable to those of wild-type mice (Fig. 4c), suggesting a critical role for RIP1 signalling in *Ptpn6^{spn}*-induced inflammatory disease. RIP1-deficient mice suffer from perinatal lethality²¹, hampering genetic analysis of the role of RIP1 in *Ptpn6^{spn}*-induced autoinflammation. However, the observation that *Ptpn6^{spn}*-mediated autoinflammation stems from the haematopoietic compartment (Fig. 4a, b) provided a rationale to explore the role of RIP1 by means of fetal liver transplantation experiments. To this end, fetal liver cells

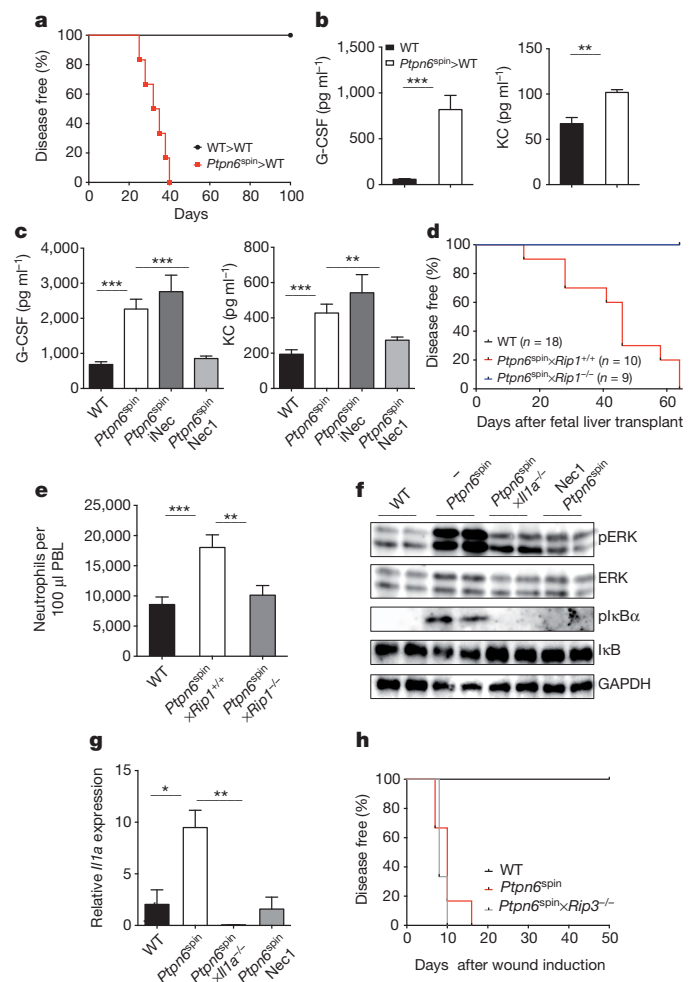


Figure 4 | RIP1 regulates *Ptpn6^{spn}*-mediated disease through the control of proinflammatory signalling and *Il1a* expression and not via RIP3-induced necroptosis. **a**, Spontaneous incidence of footpad inflammation in bone marrow chimaeric mice (donor > recipient). **b**, Levels of circulating granulopoiesis-associated factors in bone marrow chimaera. **c**, Wild-type mice were pre-treated with vehicle control ($n = 27$) and *Ptpn6^{spn}* mice were pre-treated with vehicle control ($n = 22$), 50 μ g necrostatin 1 (Nec1) ($n = 33$), or 50 μ g of an inactive control analogue of Nec1 (iNec) ($n = 10$) for 1 h before microabrasion injury induction. Serum levels of granulopoiesis-inducing factors 4 h after wound induction. **d**, **e**, Spontaneous incidence of footpad inflammation (**d**) and numbers of peripheral blood neutrophils (**e**) in wild-type (*Ptpn6^{WT} × Rip1^{+/+}* > WT), *Ptpn6^{spn} × Rip1^{+/+}* (*Ptpn6^{spn} × Rip1^{+/+}* > WT) and *Ptpn6^{spn} × Rip1^{-/-}* (*Ptpn6^{spn} × Rip1^{-/-}* > WT) fetal liver transplant mice. **f**, **g**, Wild-type, *Ptpn6^{spn}* and *Ptpn6^{spn} × Il1a^{-/-}* mice were pre-treated with PBS or 50 μ g Nec1 1 h before microabrasion injury induction. Regulation of ERK and NF- κ B signalling (**f**) and *Il1a* expression (**g**) in the footpads 2 h after wound induction. **h**, Microabrasion injuries were induced on the plantar surfaces of the footpads of wild-type, *Ptpn6^{spn}* and *Ptpn6^{spn} × Rip3^{-/-}* mice. Percentages of disease-free mice over time. Data show mean \pm s.e.m. of a representative experiment. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

from *Ptpn6*^{WT}*Rip1*^{+/+}, *Ptpn6*^{spn}*Rip1*^{+/+} and *Ptpn6*^{spn}*Rip1*^{-/-} embryos collected at embryonic (E) day E14.5 were transferred into irradiated CD45.1 congenic mice. Reconstitution of recipient mice with control *Ptpn6*^{spn}*Rip1*^{+/+} fetal liver cells resulted in unremitting footpad swelling, whereas genetic deletion of *Rip1* in the haematopoietic compartment provided protection against *Ptpn6*^{spn}-associated inflammatory disease progression (Fig. 4d and Supplementary Fig. 17) and neutrophilia (Fig. 4e), a hallmark of this inflammatory syndrome. We proposed that targeted MAP kinase and NF- κ B signalling drives *Ptpn6*^{spn}-associated inflammation. In agreement, we found that *in vivo* RIP1 inhibition markedly dampened local activation of ERK and NF- κ B signalling (Fig. 4f). Moreover, pharmacological blockade of NF- κ B activation with the IKK- β inhibitor SC-514 and inhibition of ERK signalling with U0126 treatment both abrogated hyperinflammatory cytokine production in *Ptpn6*^{spn} mice (Supplementary Fig. 18). Importantly, the RIP1 kinase inhibitor Nec1 also inhibited the synthesis of *Il1a* transcripts (Fig. 4g), further highlighting the role of RIP1 as a critical regulator of NF- κ B-induced and IL-1 α -driven autoinflammation in *Ptpn6*^{spn} mice. Notably, IL-1 α deletion also attenuated exacerbated ERK and NF- κ B signalling in the footpads of *Ptpn6*^{spn} mice (Fig. 4f), suggesting that RIP1-mediated IL-1 α production triggers an inflammatory feedback loop that contributes to disease progression. In addition to driving MAP kinase and NF- κ B activation, RIP1 controls induction of necroptosis in conjunction with RIP3 (ref. 20). To verify a potential role for unwarranted necroptosis induction in *Ptpn6*^{spn}-associated inflammatory disease, *Rip3*-deficient mice were bred to *Ptpn6*^{spn} mice. However, unlike deletion of *Rip1* and *Il1a*, genetic ablation of *Rip3* expression failed to protect *Ptpn6*^{spn} mice from exacerbated inflammation in response to microabrasion-induced tissue injury (Fig. 4h and Supplementary Fig. 19). These results indicate that RIP3-mediated necroptosis is dispensable, and suggest a critical role for RIP1-mediated regulation of MAP kinase and NF- κ B signalling in driving the inflammatory phenotype of *Ptpn6*^{spn} mice.

Defective neutrophil homeostasis is associated with numerous devastating human diseases²³. For instance, neutropenia can cause severe susceptibility to infection, whereas neutrophilia is linked to autoinflammatory disorders. Our results in the *Ptpn6*^{spn} mouse inflammation model highlight a critical role for RIP1-mediated ERK and NF- κ B signalling in haematopoietic cells in driving an inflammatory circuit that triggers excessive inflammatory responses and persistent tissue damage. Indeed, biochemical and genetic blockade of RIP1 signalling prevented inflammatory cytokine production and protected *Ptpn6*^{spn} mice from autoinflammation. IL-1 α was critical for RIP1-mediated inflammatory disease progression, which proceeded independently of inflammasome/caspase-1-produced IL-1 β and RIP3-mediated necroptosis. Consequently, therapeutic inhibition of RIP1 activity and/or neutralization of IL-1 α may provide novel approaches to break the self-reinforcing inflammatory circuits that drive chronic autoinflammatory and autoimmune diseases.

METHODS SUMMARY

Ptpn6^{spn} mice homozygous for the Tyr208Asn amino acid substitution in the C-terminal Src homology 2 domain of SHP-1 have been described previously⁹. *Ptpn6*^{spn} mice spontaneously develop a persistent footpad disease that is characterized by paw swelling and cutaneous inflammation at 8–16 weeks of age. Blood was collected by submandibular venipuncture to measure the levels of circulating neutrophils and inflammatory cytokines. To assess T-cell-mediated cytokine production, splenocytes and popliteal lymph node cells were re-stimulated with PMA/ionomycin followed by intracellular flow cytometry staining. Formalin-preserved footpad samples were embedded in paraffin. Footpad infiltration by inflammatory cells and neutrophils was assessed in a blinded manner by a pathologist using haematoxylin and eosin staining and neutrophil immunohistochemistry. The accumulation of immune cells in lymphoid organs was evaluated with the use of flow cytometry staining. To provoke microabrasion injury, mice were anaesthetized and the plantar surfaces of their hind paws were irritated by gently rubbing with sterile sandpaper. Clinical scores were assigned based on oedema, erythema and weepy wound formation. The development of persistent footpad swelling was used to evaluate disease incidence over time. EAE was induced using

MOG peptide, CFA, heat-inactivated *Mycobacterium tuberculosis* and *Bordetella pertussis* toxin. Neutrophils were purified from the bone marrow and stimulated with LPS. Fetal liver cells collected at embryonic day E14.5 were transplanted into lethally irradiated wild-type mice to generate *Ptpn6*^{spn} \times *Rip1*^{-/-} wild-type chimaeric mice.

Full Methods and any associated references are available in the online version of the paper.

Received 16 June 2012; accepted 9 April 2013.

Published online 26 May 2013.

- Pao, L. I., Badour, K., Siminovich, K. A. & Neel, B. G. Nonreceptor protein-tyrosine phosphatases in immune cell signaling. *Annu. Rev. Immunol.* **25**, 473–523 (2007).
- Nesterovitch, A. B. *et al.* Spontaneous insertion of a b2 element in the *ptpn6* gene drives a systemic autoinflammatory disease in mice resembling neutrophilic dermatosis in humans. *Am. J. Pathol.* **178**, 1701–1714 (2011).
- Crocker, B. A. *et al.* Inflammation and autoimmunity caused by a SHP1 mutation depend on IL-1, MyD88, and a microbial trigger. *Proc. Natl Acad. Sci. USA* **105**, 15028–15033 (2008).
- Cao, H. & Hegele, R. A. Identification of polymorphisms in the human *SHP1* gene. *J. Hum. Genet.* **47**, 445–447 (2002).
- Christophi, G. P. *et al.* SHP-1 deficiency and increased inflammatory gene expression in PBMCs of multiple sclerosis patients. *Lab. Invest.* **88**, 243–255 (2008).
- Green, M. C. & Shultz, L. D. Motheaten, an immunodeficient mutant of the mouse. I. Genetics and pathology. *J. Hered.* **66**, 250–258 (1975).
- Shultz, L. D., Coman, D. R., Bailey, C. L., Beamer, W. G. & Sidman, C. L. “Viable motheaten,” a new allele at the motheaten locus. I. Pathology. *Am. J. Pathol.* **116**, 179–192 (1984).
- Shultz, L. D. *et al.* Mutations at the murine motheaten locus are within the hematopoietic cell protein-tyrosine phosphatase (*Hcph*) gene. *Cell* **73**, 1445–1454 (1993).
- Tsui, H. W., Siminovich, K. A., de Souza, L. & Tsui, F. W. Motheaten and viable motheaten mice have mutations in the haematopoietic cell phosphatase gene. *Nature Genet.* **4**, 124–129 (1993).
- Lorenz, U. SHP-1 and SHP-2 in T cells: two phosphatases functioning at many levels. *Immunol. Rev.* **228**, 342–359 (2009).
- Lukens, J. R., Dixit, V. D. & Kanneganti, T. D. Inflammasome activation in obesity-related inflammatory diseases and autoimmunity. *Discov. Med.* **12**, 65–74 (2011).
- Tschopp, J. & Schroder, K. NLRP3 inflammasome activation: The convergence of multiple signalling pathways on ROS production? *Nature Rev. Immunol.* **10**, 210–215 (2010).
- Chen, C. J. *et al.* Identification of a key pathway required for the sterile inflammatory response triggered by dying cells. *Nature Med.* **13**, 851–856 (2007).
- Cohen, I. *et al.* Differential release of chromatin-bound IL-1 α discriminates between necrotic and apoptotic cell death by the ability to induce sterile inflammation. *Proc. Natl Acad. Sci. USA* **107**, 2574–2579 (2010).
- Rider, P. *et al.* IL-1 α and IL-1 β recruit different myeloid cells and promote different stages of sterile inflammation. *J. Immunol.* **187**, 4835–4843 (2011).
- Kranich, J., Maslowski, K. M. & Mackay, C. R. Commensal flora and the regulation of inflammatory and autoimmune responses. *Semin. Immunol.* **23**, 139–145 (2011).
- Henaoui-Mejia, J. *et al.* Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* **482**, 179–185 (2012).
- Kubes, P. & Mehal, W. Z. Sterile inflammation in the liver. *Gastroenterology* **143**, 1158–1172 (2012).
- Chen, G. Y. & Nunez, G. Sterile inflammation: sensing and reacting to damage. *Nature Rev. Immunol.* **10**, 826–837 (2010).
- Vandenabeele, P., Declercq, W., Van Herreweghe, F. & Vanden Berghe, T. The role of the kinases RIP1 and RIP3 in TNF-induced necrosis. *Sci. Signal.* **3**, re4 (2010).
- Kelliher, M. A. *et al.* The death domain kinase RIP mediates the TNF-induced NF- κ B signal. *Immunity* **8**, 297–303 (1998).
- Takahashi, N. *et al.* Necrostatin-1 analogues: critical issues on the specificity, activity and *in vivo* use in experimental disease models. *Cell Death Dis.* **3**, e437 (2012).
- Amulic, B., Cazalet, C., Hayes, G. L., Metzler, K. D. & Zychlinsky, A. Neutrophil function: from mechanisms to disease. *Annu. Rev. Immunol.* **30**, 459–489 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. A. Buettner, V. M. Dixit and D. R. Green for supply of mutant mice. We thank R. Weinlich for helpful discussions. We thank M. Johnson in the St Jude Small Animal Imaging Center for helping to evaluate embryonic development in fetal liver transplantation studies. M.L. is supported by European Union Marie-Curie grant 256432, ERC grant 281600, and grants G030212N, 1.2.201.10.N.00 and 1.5.122.11.N.00 from the Fund for Scientific Research-Flanders. This work was supported by grants from the National Institutes of Health (grants AR056296, CA163507 and AI101935) and the American Lebanese Syrian Associated Charities (ALSAC) to T.-D.K.

Author Contributions J.R.L., M.L. and T.-D.K. designed the study; J.R.L. performed experiments and G.R.J. provided technical assistance; J.R.L., M.L. and T.-D.K. analysed data and wrote the manuscript; M.A.K. and Y.I. provided genetic tools; P.V. performed and analysed the histopathology data; T.-D.K. oversaw the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.-D.K. (Thirumala-Devi.Kanneganti@StJude.org).

METHODS

Mice. *Ptpn6*^{pin} (ref. 3), *Nlrp3*^{-/-} (ref. 24), *Casp1*^{-/-} (ref. 24), *Il1b*^{-/-} (ref. 25), *Il1a*^{-/-} (ref. 26), *Rip3*^{-/-} (ref. 27) and *Rip1*^{-/-} (ref. 21) mice were previously described. All mice were housed under specific pathogen-free conditions within the Animal Resource Center at St Jude Children's Research Hospital. Animal studies were conducted under protocols approved by the Institutional Animal Care and Use Committee of St Jude Children's Research Hospital.

Histopathology. Formalin-preserved feet were processed and embedded in paraffin according to standard procedures. Sections (5 µm) were stained with haematoxylin and eosin (H&E) and examined by a pathologist blinded to the experimental groups. For immunohistochemistry, formalin-fixed paraffin-embedded tissues were cut into 4 µm sections and slides were stained with anti-Gr-1 to stain neutrophils in the footpads.

Microabrasion injury model. A novel microabrasion wound model was developed to evaluate the inflammatory response in a synchronized and controlled fashion. In this model, wild-type and disease-free *Ptpn6*^{pin} mice (4–8 weeks old) were anaesthetized and the plantar surfaces of the hind paws were irritated by gently rubbing with sterile sandpaper to induce physical trauma and microinjuries. Clinical scores were assigned daily based on the following scale: 0, no disease; 1, erythema; 2, erythema and mild swelling; 3, erythema, swelling and crusty wound formation; 4, weepy wound formation and severe swelling. The development of persistent footpad swelling was used to evaluate disease incidence over time. The levels of proinflammatory mediators that are produced in response to microabrasion injury were measured in the serum 4–5 h after wound induction. For the *in vivo* necrostatin 1 experiments, mice were given either 50 µg necrostatin 1 (Nec1) (Sigma-Aldrich) or 50 µg of an inactive control analogue (iNec, Calbiochem) by the intraperitoneal route 1 h before microabrasion irritation of the footpads. Circulating cytokine levels were measured in the serum 4–5 h later.

Experimental autoimmune encephalomyelitis (EAE). Age- (5–8 weeks) and sex-matched mice were immunized subcutaneously with 100 µg MOG_{35–55} peptide (MEVGWYRSPFSRVVHLYRNGK) emulsified in CFA (Difco Laboratories) with 500 µg *Mycobacterium tuberculosis* on day 0. Mice also received 200 ng pertussis toxin (List Biological Laboratories) by intraperitoneal injection on days 0 and 2. Disease severity was assessed daily by assigning clinical scores according to the following scale: 0, no disease; 1, tail paralysis; 2, weakness of hindlimbs; 3, paralysis of hindlimbs; 4, paralysis of hindlimbs and severe hunched posture; 5, moribund or death. To collect CNS leukocytes, mice were perfused through the left ventricle with PBS. The spinal cord was isolated, cut into small pieces, and then passed through a 70 µm cell strainer. Leukocytes were then purified by gradient centrifugation using a 38% Percoll solution. Cells were washed once in PBS and then re-suspended in media.

In vivo serum cytokines. Blood was collected by submandibular venipuncture and allowed to clot for 30–60 min at room temperature. Serum was collected after centrifugation and cytokines were measured by ELISA.

ELISA. Cytokine ELISA was performed according to manufacturer's instructions (Millipore).

Flow cytometry and antibodies. The following monoclonal antibodies were used for flow cytometric cell marker analysis: CD4 (L3T4), IFN-γ (XMG1.2), IL-17A (eBio17B7), MHCII (M5/114.15.2), CD11b (M1/70), CD19 (6D5), CD44 (IM7), Ly-6G (1A8), CD25 (3C7), B220 (RA3-6B2) and Gr-1 (RB6-8C5) from eBioscience and TCR-β (H57-597), CD8 (53-6.7), Foxp3 (FJK-16s), CD62L (MEL-14), TNF-α (MP6-XT22), CD11c (N418), CD45.1 (A20) and CD45.2 (104) from Biolegend. Intracellular cytokine staining was done using the eBioscience IC fixation/permeabilization kit according to the manufacturer's protocol. Intracellular staining for the Foxp3 transcription factor was performed using the eBioscience Foxp3 staining set according to the manufacturer's recommendations. Flow cytometry data were acquired on an upgraded five-colour FACScan or multi-colour LSRII (BD) and were analysed with FlowJo software (TreeStar).

Ex vivo lymphocyte re-stimulation. Splenocytes and lymph node (popliteal and mesenteric) cells were collected and re-stimulated with 20 ng ml⁻¹ phorbol 12-myristate 13-acetate (PMA) and 500 ng ml⁻¹ ionomycin in the presence of monensin for 3–4 h. Cells were stained according to the manufacturer's instructions (eBioscience). For the EAE experiment, splenocytes were harvested and re-stimulated with 30 µg ml⁻¹ MOG peptide. Supernatants were collected after 48 h to measure cytokine levels by ELISA.

Bone marrow chimaeras. Bone marrow was flushed from the femurs and filtered through a 40 µm filter. 3–5 × 10⁶ cells in 200 µl PBS were transferred by tail vein injection into lethally irradiated (1,000 rad) mice. Congenic CD45 markers were used to verify chimaerism.

In vitro macrophage stimulation. Bone-marrow-derived macrophages (BMDMs) were generated by culturing bone marrow cells in L-cell-conditioned IMDM medium supplemented with 10% FBS, 1% non-essential amino acid, and 1% penicillin-streptomycin for 5 days. BMDMs were seeded in 12-well cell culture plates and cultured overnight. To evaluate cytokine production, BMDMs were primed with 2 µg ml⁻¹ ultrapure *Escherichia coli*-derived LPS (Invivogen) for 3 h followed by 5 mM ATP (Sigma-Aldrich) for an additional 30 min. BMDMs were also separately stimulated with *Salmonella enterica* serovar Typhimurium (5 MOI) for 4 h and supernatants were collected to evaluate cytokine secretion by ELISA.

Neutrophil culture and in vitro stimulation. Bone marrow cells were isolated from the femurs of mice and neutrophils (CD11b⁺ Gr-1⁺) were purified by flow cytometry sorting. Neutrophils (1 × 10⁶ cells ml⁻¹) were stimulated with 100 ng ml⁻¹ ultrapure *Escherichia coli*-derived LPS (Invivogen). Supernatants were collected after 48 h of stimulation and cytokine levels were measured by ELISA.

Western blotting. Footpad protein lysates were collected in RIPA lysis buffer supplemented with complete protease inhibitor cocktail (Roche) and PhosSTOP (Roche) using a tissue homogenizer. Samples were resolved by SDS-PAGE and transferred to polyvinylidene difluoride (PVDF) membranes via electroblotting. Membranes were blocked in 5% non-fat milk and incubated overnight at 4 °C with primary antibodies. The membranes were then probed with horseradish peroxidase (HRP)-tagged secondary antibodies at room temperature for 1 h. Immunoreactive proteins were visualized using the ECL method (Pierce).

Real-time RT-PCR. Total RNA was isolated from the hind paws with Trizol (Invitrogen) according to the manufacturer's instructions. 1 µg of RNA was reverse-transcribed to cDNA with random RNA-specific primers using the high-capacity cDNA reverse transcription kit (Applied Biosystems). Transcript levels of *Il1a* and *Gapdh* were analysed using SYBR-Green (Applied Biosystems) on an ABI7500 real-time PCR machine according to the manufacturers' recommendations. Relative expression was calculated using the ΔΔCt standardization method.

Footpad pathology scoring. Footpad haematoxylin and eosin sections were scored based on the extent and severity of inflammation, ulceration and hyperplasia of the mucosa in a blinded fashion by a veterinary pathologist. Severity scores for inflammation were as follows: 0, normal (within normal limits); 1, minimal (small, focal, or widely separated); 2, mild; 3, moderate (moderate multifocal inflammation with dermatitis, suppurative, coalescing with intraepithelial and follicular abscesses); 4, marked (marked inflammation, with intraepidermal pustules, epidermal hyperplasia, acantholysis, dermatitis, perifolliculitis); 5, severe (severe inflammation, with intraepidermal pustules, epidermal hyperplasia, acantholysis, dermatitis, perifolliculitis, lesions covering >50% of the section).

APAP-induced hepatotoxicity model. Acetaminophen (Sigma-Aldrich) was dissolved in sterile PBS by heating the solution to 55 °C. Mice that fasted overnight for 16–18 h received 250 mg kg⁻¹ of acetaminophen (APAP) by intraperitoneal injection. Mice were harvested 18–20 h after injection and the levels of serum alanine aminotransferase (sALT) were measured in the blood by ELISA.

Statistical analysis. All results are presented as means ± standard errors. We performed statistical analysis using the two-tailed Student's *t*-test. Differences were considered statistically significant when *P* < 0.05. *P* values are denoted by **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

24. Kanneganti, T. D. *et al.* Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. *Nature* **440**, 233–236 (2006).
25. Shornick, L. P. *et al.* Mice deficient in IL-1β manifest impaired contact hypersensitivity to trinitrochlorobenzene. *J. Exp. Med.* **183**, 1427–1436 (1996).
26. Matsuki, T., Nakae, S., Sudo, K., Horai, R. & Iwakura, Y. Abnormal T cell activation caused by the imbalance of the IL-1/IL-1R antagonist system is responsible for the development of experimental autoimmune encephalomyelitis. *Int. Immunol.* **18**, 399–407 (2006).
27. Newton, K., Sun, X. & Dixit, V. M. Kinase RIP3 is dispensable for normal NF-κBs, signaling by the B-cell and T-cell receptors, tumor necrosis factor receptor 1, and Toll-like receptors 2 and 4. *Mol. Cell. Biol.* **24**, 1464–1469 (2004).

Vector transmission regulates immune control of *Plasmodium* virulence

Philip J. Spence¹, William Jarra¹, Prisca Lévy¹, Adam J. Reid², Lia Chappell², Thibaut Brugat¹, Mandy Sanders², Matthew Berriman² & Jean Langhorne¹

Defining mechanisms by which *Plasmodium* virulence is regulated is central to understanding the pathogenesis of human malaria. Serial blood passage of *Plasmodium* through rodents^{1–3}, primates⁴ or humans⁵ increases parasite virulence, suggesting that vector transmission regulates *Plasmodium* virulence within the mammalian host. In agreement, disease severity can be modified by vector transmission^{6–8}, which is assumed to ‘reset’ *Plasmodium* to its original character³. However, direct evidence that vector transmission regulates *Plasmodium* virulence is lacking. Here we use mosquito transmission of serially blood passaged (SBP) *Plasmodium chabaudi chabaudi*⁹ to interrogate regulation of parasite virulence. Analysis of SBP *P. c. chabaudi* before and after mosquito transmission demonstrates that vector transmission intrinsically modifies the asexual blood-stage parasite, which in turn modifies the elicited mammalian immune response, which in turn attenuates parasite growth and associated pathology. Attenuated parasite virulence associates with modified expression of the *pir* multi-gene family. Vector transmission of *Plasmodium* therefore regulates gene expression of probable variant antigens in the erythrocytic cycle, modifies the elicited mammalian immune response, and thus regulates parasite virulence. These results place the mosquito at the centre of our efforts to dissect mechanisms of protective immunity to malaria for the development of an effective vaccine.

The definitive host for mammalian *Plasmodium* is the anopheline mosquito. Within this vector, a complex series of developmental events, including fertilization and meiosis, culminates in invasion of the salivary glands by infective sporozoites, which are transmitted to the mammalian host through mosquito bite. Sporozoites deposited in the dermis migrate to the liver, invade hepatocytes and undergo further developmental processes before the release of merozoites that invade erythrocytes. The subsequent erythrocytic cycle is entirely responsible for the morbidity and mortality associated with malaria. The complexity of the *Plasmodium* life cycle has led to much of the basic biology of the blood-stage infection being studied in isolation, with *in vivo* experiments largely initiated through direct injection of infected erythrocytes. However, serial blood passage of *Plasmodium* increases parasite virulence^{1–5}, suggesting that regulation of *Plasmodium* virulence is an inherent consequence of vector transmission³. This could result indirectly from vector control of inoculum size or the passage of large parasite populations through extreme bottlenecks, although these consequences of mosquito transmission are not thought to be major determinants of disease severity^{8,10}. Alternatively, vector transmission may regulate *Plasmodium* virulence by intrinsically modifying the parasite and its interaction with the mammalian host. In this context, the immune response elicited by *Plasmodium* influences disease severity¹¹, and can therefore dictate parasite virulence. The interrelationship between the vector, parasite and mammalian immune system could thus underpin the pathogenesis of malaria.

To study regulation of *Plasmodium* virulence we developed routine mosquito transmission of SBP *P. c. chabaudi*⁹, a rodent malaria parasite

that has many characteristics associated with the pathogenesis of human infection¹². This allowed us to directly compare SBP parasites before and after vector transmission. Accordingly, mice were infected with SBP *P. c. chabaudi* AS either by injection of parasitized erythrocytes (pE) or mosquito bite (see Methods). Following mosquito transmission, asexual blood-stage parasite growth was attenuated (Fig. 1a), and a low-grade, recrudescent infection with extended chronicity was established (Supplementary Fig. 1). Attenuated parasite growth in the erythrocytic cycle was not influenced by dose (ref. 9 and Supplementary Fig. 2) or, importantly, by the pre-erythrocytic stages of infection, as attenuated parasite growth was similarly observed when mice were injected with pE derived from recently mosquito-transmitted (MT) parasite lines (Fig. 1b). Similar results were observed with cloned parasites derived from SBP *P. c. chabaudi* AS (Supplementary Fig. 3), and with the hypervirulent *P. c. chabaudi* CB (Supplementary Fig. 4). Mosquito transmission therefore attenuated the asexual blood-stage parasite. As expected, serial blood passage of MT *P. c. chabaudi* AS rapidly increased parasite growth (Supplementary Fig. 5). Mice infected with *P. c. chabaudi* AS through mosquito bite did not show the severe hypothermia, cachexia or hepatic cellular damage that was observed during the acute phase of infection in mice injected with SBP parasites, although they still showed severe anaemia despite attenuated parasite growth (Fig. 1c–f). Mosquito transmission therefore reduced disease severity in the mammalian host. Despite attenuated parasite growth and reduced pathogenicity, MT *P. c. chabaudi* AS elicited robust, long-term protection to reinfection with homologous or heterologous blood-stage parasites (Fig. 1g and Supplementary Fig. 6). Thus, vector transmission regulates the virulence of *Plasmodium* by intrinsically modifying the asexual blood-stage parasite, without influencing the capacity of the mammalian host to acquire robust immunity to reinfection.

The pathogenesis of malaria is complex and influenced by the mammalian immune system; dysregulated immune reactions can directly promote severe disease¹¹, whereas an appropriate response can enhance parasite clearance without promoting pathology¹³. The immune response induced by *Plasmodium* can therefore define its virulence. Throughout the erythrocytic cycle the spleen is the major anatomical site associated with the developing immune response¹⁴, and mice infected with *P. c. chabaudi* AS through mosquito bite developed marked splenomegaly with rapid recruitment of inflammatory monocytes (Supplementary Figs 7 and 8). Importantly, following mosquito transmission there was enhanced expansion of activated CD8 α^+ and CD8 α^- dendritic cells, which present malaria-specific antigens and stimulate CD4 $^+$ T-cell proliferation¹⁵, in the acute phase of infection (Fig. 2a and Supplementary Fig. 9). Correspondingly, the magnitude of the effector CD4 $^+$ T-cell response, which orchestrates innate and adaptive immune control of blood-stage parasite growth¹³, was also enhanced following mosquito transmission, and the memory CD4 $^+$ T-cell population showed a predominantly effector memory phenotype (Fig. 2b and Supplementary Fig. 10). Infection with MT *P. c. chabaudi* AS also increased the magnitude of the class-switched malaria-specific antibody response, a central component of erythrocytic immunity¹¹

¹Division of Parasitology, MRC National Institute for Medical Research (NIMR), Mill Hill, London NW7 1AA, UK. ²Parasite Genomics, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.

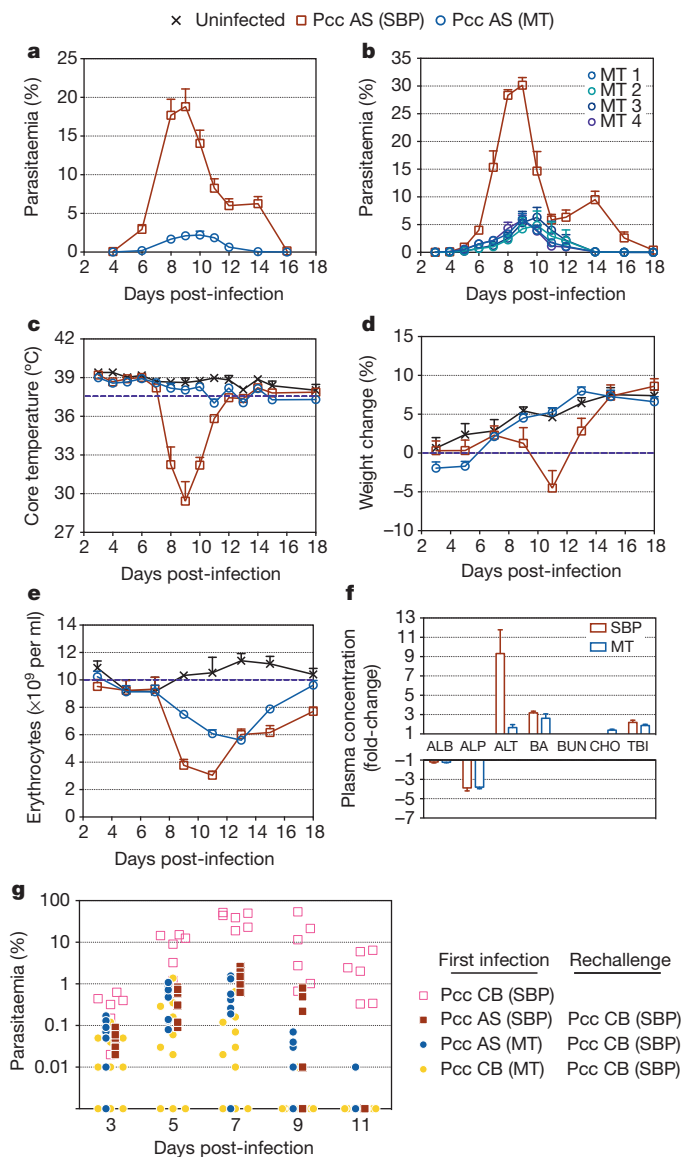


Figure 1 | Mosquito transmission of *P. c. chabaudi* AS attenuates virulence. **a**, Parasitaemia of C57BL/6 mice injected with 10^5 SBP *P. c. chabaudi* AS (Pcc AS) or infected with Pcc AS through mosquito bite. **b**, Parasitaemia of C57BL/6 mice injected with 10^5 SBP Pcc AS or injected with 10^4 or 10^5 pE derived from one of four recently MT lines of Pcc AS. **c–e**, Temperature (**c**), weight (**d**) and erythrocyte count (**e**) of C57BL/6 mice injected with 10^5 SBP Pcc AS or infected with Pcc AS through mosquito bite. **f**, Liver enzyme concentration on day 10 post-infection in plasma of C57BL/6 mice injected with 10^5 SBP Pcc AS or infected with Pcc AS through mosquito bite. Data presented as fold-change relative to uninfected control mice. ALB, albumin; ALP, alkaline phosphatase; ALT, alanine aminotransferase; BA, bile acids; BUN, blood urea nitrogen; CHO, cholesterol; TBI, total bilirubin. **g**, Parasitaemia of C57BL/6 mice injected with 10^6 SBP *P. c. chabaudi* CB (Pcc CB) as a first infection (open symbols), or as a rechallenge (closed symbols) 90 days after injection with 10^5 SBP Pcc AS or infection with Pcc AS or CB through mosquito bite. ($n = 3–20$ mice per group; data presented as mean with s.e.m.).

(Fig. 2c, d). Mosquito transmission therefore enhanced antibody production in the chronic phase of infection, subsequent to enhanced innate and adaptive cellular responses early in infection. Conversely, mosquito transmission attenuated systemic inflammation during the acute phase response, with decreased circulating levels of pro-inflammatory cytokines and chemokines associated with severe disease^{11,16} (Fig. 2e). Thus, vector transmission intrinsically modifies the asexual blood-stage parasite and transforms the mammalian immune response elicited during the erythrocytic cycle.

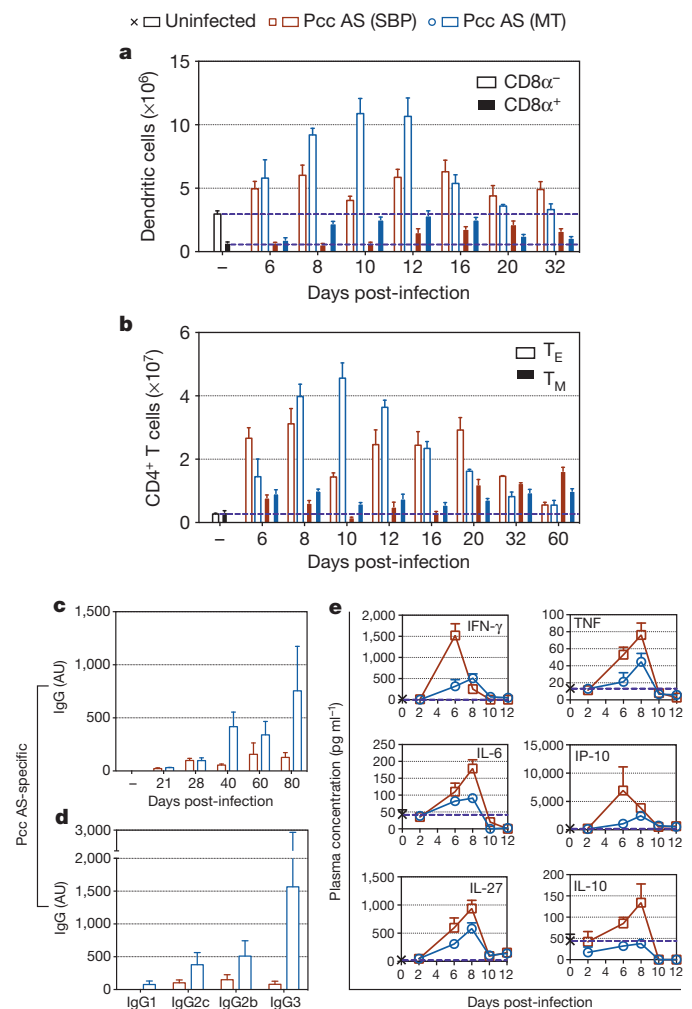


Figure 2 | Mosquito transmission of *P. c. chabaudi* AS transforms the elicited mammalian immune response. **a**, **b**, Number of CD8 α^+ (open bars) and CD8 α^+ (closed bars) dendritic cells (**a**), and effector (T_E) (open bars) and memory (T_M) (closed bars) CD4⁺ T cells (**b**) in spleens of C57BL/6 mice injected with 10^5 SBP Pcc AS or infected with Pcc AS through mosquito bite. **c**, **d**, Plasma concentration of total parasite-specific IgG throughout infection (**c**) and parasite-specific IgG subclasses on day 80 post-infection (**d**) in C57BL/6 mice injected with 10^5 SBP Pcc AS or infected with Pcc AS through mosquito bite. Data presented as arbitrary units (AU) relative to hyper-immune plasma. **e**, Plasma cytokine concentration in C57BL/6 mice injected with 10^5 SBP Pcc AS or infected with Pcc AS through mosquito bite. ($n = 3–5$ mice per group per time-point; data presented as mean with s.e.m.).

Parasite growth and pathogenicity are, in part, determined by host susceptibility. Infection of susceptible mouse strains with *P. c. chabaudi* AS through mosquito bite causes severe disease and death⁹, demonstrating that vector transmission does not limit the potential virulence of the asexual blood-stage parasite. We therefore addressed whether attenuated parasite virulence in C57BL/6 mice infected with MT *P. c. chabaudi* AS was a consequence of the transformed host immune response. Immunodeficient mice were injected with SBP *P. c. chabaudi* AS, or with an equivalent number of pE derived from a recently MT line. Disruption of the innate and adaptive immune responses, through depletion of CD4⁺ T cells, or depletion of the entire adaptive arm of the immune system led to a virulent and fatal acute phase infection (Fig. 3). Attenuation of parasite virulence by mosquito transmission was therefore dependent upon an intact mammalian immune response and, furthermore, independent of parasite growth rate (Supplementary Fig. 11). Immune control of parasite virulence therefore resulted directly and exclusively from modified innate and

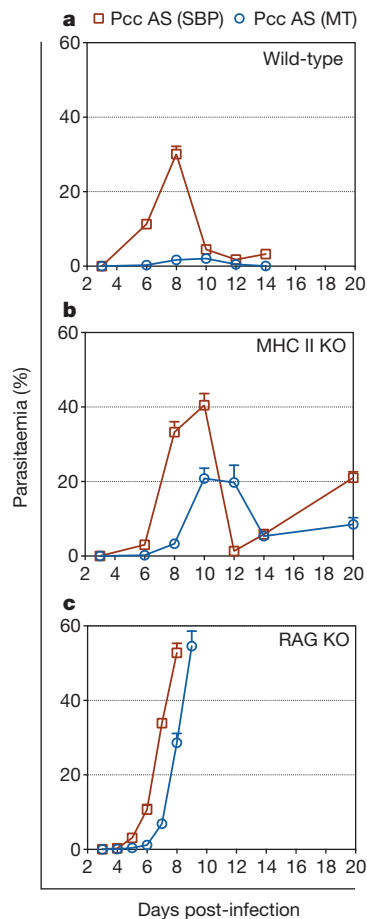


Figure 3 | Transformed innate and adaptive immune responses attenuate *P. c. chabaudi* AS virulence. a–c, Parasitaemia of wild-type (a), CD4⁺ T-cell deficient (MHC II KO) (b), and B- and T-cell deficient (RAG KO) (c) C57BL/6 mice injected with 10^5 SBP Pcc AS or injected with 10^5 pE derived from a recently MT line of Pcc AS. ($n = 4–6$ mice per group; data presented as mean with s.e.m.).

adaptive immune responses elicited by, and directed against, the blood-stage parasite. Vector transmission of *Plasmodium* thus intrinsically modifies the asexual blood-stage parasite, which in turn modifies the

elicited mammalian immune response, which in turn regulates parasite virulence.

Defining parasite gene expression in the erythrocytic cycle after vector transmission is thus central to understanding the pathogenesis of malaria. We therefore performed genome-wide RNA sequencing on *P. c. chabaudi* AS, directly comparing blood-stage parasites before and after mosquito transmission. This allowed us to identify a set of *Plasmodium* virulence genes that direct the elicited mammalian immune response (Fig. 4). Vector transmission modified expression of approximately 10% of the entire genome in the late trophozoite stage parasite (Supplementary Tables). The majority of genes upregulated following mosquito transmission encoded exported proteins, with the potential to access and modulate the mammalian immune system. Importantly, parasite gene expression was most intensely regulated within the sub-telomeric large multi-gene families, with preferential regulation of the *pir* multi-gene family (termed *cir* in *P. c. chabaudi*) (Supplementary Fig. 12). Out of 200 *cir* genes, 123 (61.5%) were differentially expressed following mosquito transmission, with 114 *cir* genes (57%) upregulated. Furthermore, the most upregulated gene following serial blood passage was identified as the most highly expressed *cir* gene (*PCHAS_110030*) in mice infected with SBP *P. c. chabaudi* AS (ref. 17, Fig. 4 and Supplementary Fig. 12). Serial blood passage therefore selected for dominant *cir* gene expression, whereas mosquito transmission revoked the selected expression hierarchy and promoted a generalized increase in *cir* expression across the parasite population. We therefore uncover a direct association between *pir* gene expression and *Plasmodium* virulence, and demonstrate that vector transmission regulates expression of probable antigenic variants¹⁸, as proposed previously^{19,20}. Vector transmission of *Plasmodium* thus regulates parasite gene expression in the erythrocytic cycle and, consequently, regulates immune control of *Plasmodium* virulence.

Vector transmission will inherently regulate *Plasmodium* virulence within the mammalian host. Recombination of distinct parasite genotypes within the mosquito is likely to be fundamental for the evolution of virulence²¹. The results of this study reveal that vector transmission also regulates *Plasmodium* virulence by modifying parasite gene expression, and therefore the mammalian immune response, in the erythrocytic cycle. This is probably the outcome of a combination of distinct regulatory processes acting at multiple stages of the parasite life cycle, in both the mosquito vector and the mammalian host. It is therefore important to delineate the timing and mechanism(s) of regulation of parasite gene expression, in the context of the complete *Plasmodium* life cycle, to understand the molecular regulation of parasite

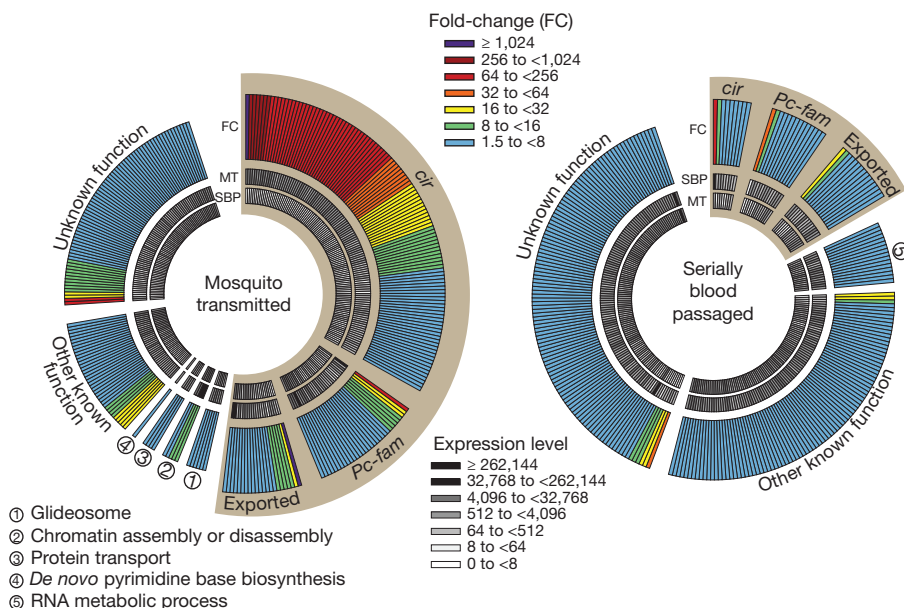


Figure 4 | Mosquito transmission of *P. c. chabaudi* AS modifies parasite gene expression in the erythrocytic cycle. C57BL/6 mice were injected with 10^5 SBP Pcc AS or infected with Pcc AS through mosquito bite. Parasites were isolated after six cycles of the blood-stage infection, and at the late trophozoite stage of development (98.3% (0.76%) and 97.0% (1.41%) trophozoites for SBP and MT samples, respectively (mean with s.d.)). Total parasite RNA was extracted and sequenced. Those genes differentially expressed between SBP and MT Pcc AS were determined; genes identified as significantly upregulated in blood-stage parasites following mosquito transmission (left) versus serial blood passage (right) are shown. Each segment represents one gene, and genes are categorised according to the function of their product and ranked based on fold-change (outer circle). The DESeq-normalized expression levels for each gene are also shown (inner circles). Sepia wedges highlight genes whose products are predicted to be exported, or otherwise accessible to the mammalian immune system.

virulence. Attenuation of parasite virulence following mosquito transmission associates with modified expression of the *pir* multi-gene family, which is conserved from rodent to human *Plasmodium*¹⁷. Importantly, vector transmission of cultured *Plasmodium falciparum* similarly modifies the composition and frequency of *var* gene expression²². Regulation of antigenic variants by vector transmission is therefore universal^{19,20,22}, and vector transmission will therefore universally regulate immune control of *Plasmodium* virulence. The interrelationship between the vector, parasite and mammalian immune system thus underpins the pathogenesis of malaria.

METHODS SUMMARY

P. c. chabaudi AS and CB were mosquito transmitted (MT) and cloned at the University of Edinburgh, UK, and sent to NIMR in 1978 and 1982, respectively. Parasites were serially blood passaged (SBP) through mice 26–32 times before use in this study. To initiate infections with SBP parasites, mice were injected intraperitoneally (i.p.) or intravenously (i.v.) with 10^4 – 10^5 pE derived from cryopreserved stocks. Alternatively, SBP parasites were transmitted through *Anopheles stephensi* and mice were infected by mosquito bite, with an estimated 9.15 infective bites per mouse⁹. We therefore directly compared SBP parasites before and after mosquito transmission. The first 52 h of infection initiated by mosquito bite was required to complete the pre-erythrocytic stages¹²; the erythrocytic cycle thus started on day 2 post-infection. To bypass the pre-erythrocytic stages, and control the dose initiating the blood-stage infection, mice were injected i.p. or i.v. with 10^4 – 10^5 pE derived from recently MT parasite lines that were just one blood passage from mosquito transmission (unless otherwise stated). The course of infection was monitored on thin blood smears by enumerating the percentage of erythrocytes infected with asexual parasites (parasitaemia). The limit of detection for patent parasitaemia was 0.01% infected erythrocytes. To determine chronicity of infection, 100 μ l blood was sub-inoculated into RAG KO mice; absence of parasitaemia in recipient mice after 14 days indicated clearance of infection in donor mice. Rechallenge studies were initiated \geq 90 days after the first infection, when \geq 95% C57BL/6 mice had naturally cleared blood-stage parasites.

Full Methods and any associated references are available in the online version of the paper.

Received 12 December 2012; accepted 26 April 2013.

Published online 29 May 2013.

1. Dearsly, A. L., Sinden, R. E. & Self, I. A. Sexual development in malarial parasites: gametocyte production, fertility and infectivity to the mosquito vector. *Parasitology* **100**, 359–368 (1990).
2. Mackinnon, M. J. & Read, A. F. Selection for high and low virulence in the malaria parasite *Plasmodium chabaudi*. *Proc. R. Soc. Lond. B* **266**, 741–748 (1999).
3. Yoeli, M., Hargreaves, B., Carter, R. & Walliker, D. Sudden increase in virulence in a strain of *Plasmodium berghei yoelii*. *Ann. Trop. Med. Parasitol.* **69**, 173–178 (1975).
4. Hartley, E. G. Increased virulence of *Plasmodium cynomolgi bastianellii* in the rhesus monkey. *Trans. R. Soc. Trop. Med. Hyg.* **63**, 411–412 (1969).
5. Chin, W., Contacos, P. G., Collins, W. E., Jeter, M. H. & Alpert, E. Experimental mosquito-transmission of *Plasmodium knowlesi* to man and monkey. *Am. J. Trop. Med. Hyg.* **17**, 355–358 (1968).

6. Alger, N. E., Branton, M., Harant, J. & Silverman, P. H. *Plasmodium berghei* NK65 in the inbred A-J mouse: variations in virulence of *P. berghei* demes. *J. Protozool.* **18**, 598–601 (1971).
7. Knowles, G. & Walliker, D. Variable expression of virulence in the rodent malaria parasite *Plasmodium yoelii yoelii*. *Parasitology* **81**, 211–219 (1980).
8. Mackinnon, M. J., Bell, A. & Read, A. F. The effects of mosquito transmission and population bottlenecks on virulence, multiplication rate and rosetting in rodent malaria. *Int. J. Parasitol.* **35**, 145–153 (2005).
9. Spence, P. J., Jarra, W., Lévy, P., Nahrendorf, W. & Langhorne, J. Mosquito transmission of the rodent malaria parasite *Plasmodium chabaudi*. *Malar. J.* **11**, 407 (2012).
10. Glynn, J. R., Collins, W. E., Jeffery, G. M. & Bradley, D. J. Infecting dose and severity of falciparum malaria. *Trans. R. Soc. Trop. Med. Hyg.* **89**, 281–283 (1995).
11. Langhorne, J., Ndungu, F. M., Sponaas, A. M. & Marsh, K. Immunity to malaria: more questions than answers. *Nature Immunol.* **9**, 725–732 (2008).
12. Stephens, R., Culleton, R. L. & Lamb, T. J. The contribution of *Plasmodium chabaudi* to our understanding of malaria. *Trends Parasitol.* **28**, 73–82 (2012).
13. Spence, P. J. & Langhorne, J. T cell control of malaria pathogenesis. *Curr. Opin. Immunol.* **24**, 444–448 (2012).
14. del Portillo, H. A. *et al.* The role of the spleen in malaria. *Cell. Microbiol.* **14**, 343–355 (2012).
15. Sponaas, A. M. *et al.* Malaria infection changes the ability of splenic dendritic cell populations to stimulate antigen-specific T cells. *J. Exp. Med.* **203**, 1427–1433 (2006).
16. Jain, V. *et al.* Plasma IP-10, apoptotic and angiogenic factors associated with fatal cerebral malaria in India. *Malar. J.* **7**, 83 (2008).
17. Lawton, J. *et al.* Characterization and gene expression analysis of the *cir* multi-gene family of *Plasmodium chabaudi chabaudi* (AS). *BMC Genomics* **13**, 125 (2012).
18. Cunningham, D., Lawton, J., Jarra, W., Preiser, P. & Langhorne, J. The *pir* multigene family of *Plasmodium*: antigenic variation and beyond. *Mol. Biochem. Parasitol.* **170**, 65–73 (2010).
19. Brannan, L. R., McLean, S. A. & Phillips, R. S. Antigenic variants of *Plasmodium chabaudi chabaudi* AS and the effects of mosquito transmission. *Parasite Immunol.* **15**, 135–141 (1993).
20. McLean, S. A., Phillips, R. S., Pearson, C. D. & Walliker, D. The effect of mosquito transmission of antigenic variants of *Plasmodium chabaudi*. *Parasitology* **94**, 443–449 (1987).
21. Manske, M. *et al.* Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).
22. Peters, J. *et al.* High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. *Proc. Natl Acad. Sci. USA* **99**, 10689–10694 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Medical Research Council (U117584248) and the Wellcome Trust (089553 and 098051). P.J.S. is the recipient of a Leverhulme Trust Early Career Fellowship. The authors thank R. Sinden, K. Baker and M. Tunnicliffe for provision of *Anopheles stephensi*, and Biological Services at NIMR. M. Blackman, G. Kassiotis and G. Stockinger are thanked for critical reading of the manuscript.

Author Contributions P.J.S., W.J. and J.L. designed the study. P.J.S., W.J., P.L., L.C. and T.B. performed the experiments. P.J.S. and A.J.R. analysed the data. M.S. and M.B. provided project management. P.J.S. wrote the manuscript.

Author Information RNA-seq datasets have been deposited in ArrayExpress with accession number E-ERAD-95. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.L. (jlango@nimr.mrc.ac.uk).

METHODS

Mice. Inbred wild-type, major histocompatibility complex class II knockout (MHC II KO)²³ and recombination activation gene 1 knockout (RAG KO)²⁴ C57BL/6 mice were bred under specific pathogen-free conditions at NIMR. All experiments were performed in accordance with UK Home Office regulations (PPL 80/2358) and approved by the ethical review panel at NIMR. Experimental mice were age- and sex-matched, housed under reverse light conditions (light 19:00–07:00, dark 07:00–19:00) at 20–22 °C, and had continuous access to mouse breeder diet and water. Measurements of clinical pathology were taken at 16:00. Core body temperature was measured with a rectal thermometer; body weight was calculated relative to a baseline measurement taken on day –2; and erythrocyte density was determined on a VetScan HMII haematology system (Abaxis). To measure liver enzymes, plasma was analysed on a VetScan Chemistry Analyzer, using a Mammalian Liver Profile reagent rotor (Abaxis).

Enumeration of blood-stage parasites by real-time PCR. Whole blood was isolated 20 h after liver merozoite egress, when parasites were at the late trophozoite stage of development and within the first cycle of schizogony. Total RNA was extracted by acid guanidinium thiocyanate-phenol-chloroform extraction²⁵, and reverse transcribed by PCR at 42 °C using 75 U MuLV reverse transcriptase and 2.5 µM random hexamer primers (both Applied Biosystems) per sample. Parasites were quantified by real-time PCR, comparing *P. c. chabaudi* AS 18S ribosomal RNA copy number between samples and a standard curve of pE prepared at the late trophozoite stage of development. The reaction mix contained TaqMan Universal PCR Master Mix (Applied Biosystems), 300 nM forward primer (5'-AAGCATTAAATAAGCGAATACATCCTTAT-3'), 300 nM reverse primer (5'-GGGAGTTTGGTTTTGACGTTTATGCG-3') and 50 nM probe (5'-6FAM-CAATTGGTTTACCTTTTGTCTTT-TAM-3'). Real-time PCR amplification was performed on an ABI Prism 7000 Sequence Detection System (Applied Biosystems), with a temperature profile as follows: 50 °C for 2 min, followed by 95 °C for 10 min, and then 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Parasite numbers were determined per 100 µl whole blood; total circulating parasites were then calculated for each mouse based on their weight and, therefore, their estimated circulating blood volume.

Blood cytometry. Single-cell suspensions of splenocytes were prepared, erythrocytes lysed, and cells enumerated on a haemocytometer. Cells were stained with monoclonal antibodies (CD3ε biotin or PerCP-Cy5.5 (145-2C11); CD4 Pacific Blue (RM4-5); CD8α Pacific Blue (53-6.7); CD11b Pacific Blue (M1/70); CD11c APC (N418); CD44 FITC (IM7); I-A^b FITC (AF6-120.1); Ly-6G PE (1A8); NK-1.1 biotin (PK136); TER-119 biotin (all from BioLegend)) (CD19 biotin (1D3); CD62L APC (MEL-14); CD127 PE (A7R34) (all from eBioscience)) (Ly-6C Alexa Fluor 647 (ER-MP20) (from AbD Serotec)) or irrelevant isotype-matched monoclonal antibodies as negative controls. PerCP/Cy5.5-conjugated streptavidin (BD Biosciences) was used secondary to biotinylated antibodies. For phenotypic analysis, samples were acquired on a CyAn (Beckman Coulter), and data were analysed with FlowJo software (TreeStar).

Antibodies and cytokines. Malaria-specific antibodies were measured in plasma by enzyme-linked immunosorbent assay. 96-well PolySorp plates (Nunc) were coated with 50 µg ml⁻¹ parasite lysate prepared from pE isolated from C57BL/6 mice infected with SBP *P. c. chabaudi* AS; twofold serial dilutions of plasma from uninfected and hyper-immune mice were used as negative and positive controls, respectively, for experimental samples; alkaline phosphatase-conjugated goat anti-mouse IgG, IgG1, IgG2c, IgG2b and IgG3 (all from SouthernBiotech) were used for detection. Samples were developed with 1 mg ml⁻¹ 4-nitrophenyl phosphate disodium salt hexahydrate (Sigma) and attenuation was measured at 405 nm. Antibody concentrations are presented as arbitrary units (AU) relative to hyper-immune plasma. Cytokines were measured in plasma by LEGENDplex Luminex custom assay (BioLegend).

Plasmodium RNA preparation. C57BL/6 mice were infected with SBP *P. c. chabaudi* AS through injection of infected erythrocytes, or through mosquito bite. Parasites were isolated at exactly 20 h into the seventh cycle of the blood-stage infection, at the late trophozoite stage of development, as follows. Whole blood was depleted of leukocytes by Plasmodipur filtration (EuroProxima); erythrocytes were centrifuged at 400g for 10 min and lysed with 0.15% (w/v) saponin (Sigma). Samples were centrifuged at 1,000g for 5 min and washed with PBS; parasites were resuspended in TRIzol (Life Technologies) and snap-frozen on dry ice. We

prepared three biological replicates of SBP *P. c. chabaudi* AS from eight mice each, and two biological replicates of MT *P. c. chabaudi* AS from 30 mice each. RNA was extracted as described²⁶, resuspended in water and DNA removed with a TURBO DNA-free Kit (Applied Biosystems), according to the manufacturer's instructions. RNA quantity/quality was determined on an Agilent 2100 Bioanalyzer RNA 6000 Nano chip.

Amplification-free RNA-seq libraries. PolyA⁺ transcripts were selected from 10 µg total RNA using Sera-Mag Oligo(dT)-coated Magnetic Particles (Thermo Scientific). RNA was diluted with water to a volume of 130 µl and fragmented to approximately 200 nucleotides using Covaris Adaptive Focused Acoustics technology (settings: 5% duty cycle; intensity 5; 200 cycles per burst for 60 s). The RNA was ethanol-precipitated and resuspended in 10 µl water. First-strand cDNA was synthesized with Random Hexamer primers and SuperScript II Reverse Transcriptase (Life Technologies), following the manufacturer's instructions. Second-strand cDNA synthesis, end repair and dA-tailing were performed using the NEBNext mRNA library kit for Illumina (New England Biolabs), eluting in a final volume of 15 µl. Sequencing templates were prepared by mixing 15 µl cDNA, 5 µl 33 µM adaptors (based on the published adaptor²⁷ with the addition of barcode sequences; oligonucleotides supplied by Integrated DNA Technologies), 25 µl Quick Ligation buffer and 5 µl Quick DNA ligase (both from New England Biolabs) and incubating for 15 min at 25 °C. Excess adaptors were removed with two rounds of clean up with 50 µl of Agencourt AMPure XP Beads (Beckman Coulter). Final libraries were eluted in 30 µl water, visualized on an Agilent Bioanalyzer 2100 High Sensitivity DNA chip and quantified by qPCR. A pool of the five indexed libraries was sequenced on an Illumina HiSeq2000, with 100-bp paired-end reads.

Analysis of RNA expression. Paired-end RNA sequencing reads were mapped to the *P. c. chabaudi* AS reference genome (September 2012 release: ftp://ftp.sanger.ac.uk/pub/pathogens/P_chabaudi/September_2012/) using Tophat²⁸ v1.4.1, with appropriate fragment size parameters and maximum intron size 10000. Read counts per gene were calculated using in-house Perl scripts and non-uniquely mapping reads were excluded. Six genes with less than 20% unique coding sequence (kmer = 100) were excluded from the analysis (PCHAS_073210; PCHAS_083750; PCHAS_100020; PCHAS_113280; PCHAS_113290; PCHAS_130130). Differential gene expression between SBP and MT *P. c. chabaudi* AS was determined using DESeq²⁹; the three SBP *P. c. chabaudi* AS replicates were compared against the two MT *P. c. chabaudi* AS replicates to determine genes upregulated in blood-stage parasites following mosquito transmission, and vice versa to determine genes upregulated following serial blood passage. In both cases a corrected *P* value cutoff of 0.01 was applied. The resulting gene lists were categorised into 'cir' (based on published annotation¹⁷), 'Pc-fam' (based on GeneDb annotation by Ulrike Boehme at the Wellcome Trust Sanger Institute, with minor reannotation), 'Exported' (based on known biology or ExportPred³⁰ prediction), 'Other known function' and 'Unknown function'. For those genes within the category 'Other known function', we sub-categorised genes based on enriched biological process GO terms using TopGO³¹; a *P* value cutoff of 0.01 was applied. We independently added 'glideosome' as a sub-category.

23. Madsen, L. *et al.* Mice lacking all conventional MHC class II genes. *Proc. Natl Acad. Sci. USA* **96**, 10338–10343 (1999).
24. Mombaerts, P. *et al.* RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).
25. Chomczynski, P. & Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nature Protocols* **1**, 581–585 (2006).
26. Kyes, S., Pinches, R. & Newbold, C. A simple RNA analysis method shows *var* and *rif* multigene family expression patterns in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **105**, 311–315 (2000).
27. Kozarewa, I. *et al.* Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature Methods* **6**, 291–295 (2009).
28. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
29. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
30. Sargeant, T. J. *et al.* Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.* **7**, R12 (2006).
31. Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).

Negligible impact of rare autoimmune–locus coding–region variants on missing heritability

Karen A. Hunt¹, Vanisha Mistry¹, Nicholas A. Bockett¹, Tariq Ahmad², Maria Ban³, Jonathan N. Barker⁴, Jeffrey C. Barrett⁵, Hannah Blackburn⁵, Oliver Brand⁶, Oliver Burren⁷, Francesca Capon⁴, Alastair Compston³, Stephen C. L. Gough⁶, Luke Jostins⁸, Yong Kong⁹, James C. Lee¹⁰, Monkol Lek¹¹, Daniel G. MacArthur¹¹, John C. Mansfield¹², Christopher G. Mathew⁴, Charles A. Mein¹³, Muddassar Mirza⁴, Sarah Nutland⁷, Suna Onengut-Gumuscu¹⁴, Efterpi Papouli⁴, Miles Parkes¹⁰, Stephen S. Rich¹⁴, Steven Sawcer³, Jack Satsangi¹⁵, Matthew J. Simmonds⁶, Richard C. Trembath¹⁶, Neil M. Walker⁷, Eva Wozniak¹³, John A. Todd⁷, Michael A. Simpson⁴, Vincent Plagnol¹⁷ & David A. van Heel¹

Genome-wide association studies (GWAS) have identified common variants of modest-effect size at hundreds of loci for common autoimmune diseases; however, a substantial fraction of heritability remains unexplained, to which rare variants may contribute^{1,2}. To discover rare variants and test them for association with a phenotype, most studies re-sequence a small initial sample size and then genotype the discovered variants in a larger sample set^{3–5}. This approach fails to analyse a large fraction of the rare variants present in the entire sample set. Here we perform simultaneous amplicon-sequencing-based variant discovery and genotyping for coding exons of 25 GWAS risk genes in 41,911 UK residents of white European origin, comprising 24,892 subjects with six autoimmune disease phenotypes and 17,019 controls, and show that rare coding-region variants at known loci have a negligible role in common autoimmune disease susceptibility. These results do not support the rare-variant synthetic genome-wide-association hypothesis⁶ (in which unobserved rare causal variants lead to association detected at common tag variants). Many known autoimmune disease risk loci contain multiple, independently associated, common and low-frequency variants, and so genes at these loci are *a priori* stronger candidates for harbouring rare coding-region variants than other genes. Our data indicate that the missing heritability for common autoimmune diseases may not be attributable to the rare coding-region variant portion of the allelic spectrum, but perhaps, as others have proposed, may be a result of many common-variant loci of weak effect^{7–10}.

Recent large-scale human sequencing studies have revealed an abundance of rare variants (which we define as minor allele frequency (MAF) < 0.5%) and shown that these are geographically localized and are more likely to have deleterious functional consequences^{11,12}. In the largest sample size studied to date¹², 202 genes in 14,002 people were re-sequenced, and ~95% of exonic variants identified were found to be rare, with 74% observed in only one or two subjects. More broadly, across ~15,000 genes, similar findings were observed in recent exome-sequencing studies of 2,440 and 6,515 subjects^{13,14}. Importantly, these studies demonstrate that even if we had reference variation databases from a million subjects, most of the rare-variant allelic spectrum of any given sample set (for example, a case–control cohort) will be unique and only identifiable by direct re-sequencing of the entire sample set.

There are only a handful of published examples of rare coding-region variants associated with common autoimmune diseases (although many examples in familial/Mendelian immune-mediated diseases). Coding-region variants in *IFIH1* associated with type 1 diabetes (MAF in controls = 0.67–2.2%)³, *TYK2* with multiple autoimmune diseases¹⁵ and *IL23R* with inflammatory bowel disease⁵, for example, are low frequency (which we define as MAF = 0.5–5%) rather than particularly rare. In other examples, the existing evidence for association, and/or the effect sizes, are relatively weak (for example, *CARD14* and psoriasis¹⁶, *IL2RA* and *IL2RB* and rheumatoid arthritis¹⁷). The association of rare coding-region variants of *NOD2* (also known as *CARD15*) in Crohn's disease probably provides the best example, albeit three low-frequency variants comprise over 80% of all the disease-causing mutations¹⁸. Most of the studies also lose power (especially for tests in which multiple rare variants are pooled into a single analysis, for example by gene) by initially sequencing only a small sample subset rather than testing the entire rare-variant content of a large case–control sample set. We sought to improve on these methods by performing highly multiplexed sequencing of sufficiently high quality to enable direct genotyping in the entirety of a large autoimmune disease case–control collection.

We selected subjects from a single population—individuals of white Northern-European ethnicity living in the UK (Methods)—to minimize any effects of population stratification. We selected to re-sequence all RefSeq exons for 25 genes from 20 GWAS-identified risk loci showing overlap between six common autoimmune disease phenotypes (autoimmune thyroid disease, coeliac disease, Crohn's disease, psoriasis, multiple sclerosis and type 1 diabetes). All genes studied were from risk loci for at least two phenotypes, all genes had known immune system function, 18 out of 20 loci had either a single candidate immune gene or all immune genes at a locus were selected (the remaining two loci had partial transcripts of another immune gene within the 0.1 centimorgan (cM) linkage disequilibrium block), and all genes and loci were densely genotyped on the Illumina ImmunoChip (Supplementary Table 1)¹⁹. We attempted high-throughput sequencing of 52,224 samples (including positive and negative controls, and repeats). We performed extensive quality control on both samples and variant calls (Methods). The final data set comprised 41,911 phenotyped individuals (autoimmune disease cases and controls), with ImmunoChip

¹Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK. ²Peninsula College of Medicine and Dentistry, Barrack Road, Exeter EX2 5DW, UK. ³University of Cambridge, Department of Clinical Neurosciences, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. ⁴Division of Genetics and Molecular Medicine, King's College London School of Medicine, 8th Floor Tower Wing, Guy's Hospital, London SE1 9RT, UK. ⁵Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ⁶Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK. ⁷Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK. ⁸Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ⁹Department of Molecular Biophysics and Biochemistry, W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, Connecticut 06510, USA. ¹⁰Department of Medicine, University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. ¹¹Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹²Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, UK. ¹³Genome Centre, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, John Vane Science Centre, Charterhouse Square, London EC1M 6BQ, UK. ¹⁴Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia 22908-0717, USA. ¹⁵Gastrointestinal Unit, Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. ¹⁶Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK. ¹⁷University College London Genetics Institute, Gower Street, London WC1E 6BT, UK.

Table 1 | Variant types in protein-coding regions of 25 genes in 41,911 phenotyped individuals

Variant type	All variants	Rare (MAF < 0.5%)*	Novel†
Nonsynonymous SNV	1,792	1,758	1,379
Splicing SNV	86	85	65
Stopgain SNV	47	47	42
Synonymous SNV	1,024	972	674
Frameshift indels	31	31	31
Nonframeshift indels	10	10	10
Total variants	2,990	2,903	2,201
Singleton	1,602	1,598	1,411
Doubleton	470	468	378

Numbers shown are after quality-control steps. Annotation performed with GENCODE V14 gene definitions. Triallelic ($n = 124$) and quadriallelic ($n = 3$) sites (combined SNVs and indels) are shown as multiple separate variants with the appropriate annotation for each non-reference allele.

*MAF in 17,019 sequenced controls.

†Not seen in dbSNP137, or 1000 Genomes Project (April 2012 release), or NHLBI (data release ESP6500SI, with 6,503 individuals).

array genotypes available for 32,806 of these individuals (Supplementary Table 2). We discovered 4,377 variant sites across all amplicons, and the genotype call rate was 99.9989% (reference homozygote as well as non-reference genotypes) across 41,911 individuals. Of these, 2,990 variants were in protein-coding regions (including exon splice sites) of the 25 genes (Table 1 and Supplementary Table 3); 97.1% of which are rare (MAF in 17,019 controls, <0.5%); 73.6% are novel when compared with current published datasets (dbSNP137, 1000 Genomes Project, National Heart, Lung, and Blood Institute (NHLBI)) containing >6,000

individuals and 67.3% are novel compared to an unpublished data set of 25,994 exome-sequenced individuals (D. G. MacArthur, personal communication); and 68.9% were only seen in one (singleton) or two (doubleton) individuals. These proportions of novel, and rare, variants are similar to recent data from other large re-sequencing studies¹².

Our very high coverage data (99.8% of 183.4 million (site X sample) genotype calls had a read depth of ≥ 40 and 96.6% had a read depth of > 100 ; Supplementary Fig. 1) enabled stringent data filtering on call rate per sample, per variant site, and other criteria (Methods). To confirm data quality, we performed further experiments and analyses as follows: (1) we genotyped one control sample 296 times (on different 48-sample microfluidic chips), and the genotype call error rate was two non-consensus genotype calls of 1,295,581 called genotypes (0.00015%); (2) 32,806 out of 41,911 subjects also had dense ImmunoChip genotyping data at the 25 genes, and genotype concordance at 91 variant sites genotyped on both platforms was 99.994%; (3) transition/transversion (Ti/Tv) rates, a quality-control measure based on expected human mutation types, were 2.434 at coding-region variants (2.427 at singletons), 2.44 at rare (MAF < 0.5%) variants (2.437 at singletons) and 2.275 at novel variants (2.273 at singletons) (definitions in Table 1); (4) we selected all (35) nonsense single nucleotide variants (SNVs) and all (39) frameshift insertions/deletions (indels) in the ImmunoChip-genotyped samples for Sanger sequencing: two variants failed assay/PCR (polymerase chain reaction) design and there was one false-positive SNV and one false-positive indel (overall false-positive rate = 2.8%). All 70 validated SNVs and indels had the same alleles in high-throughput and Sanger-sequencing

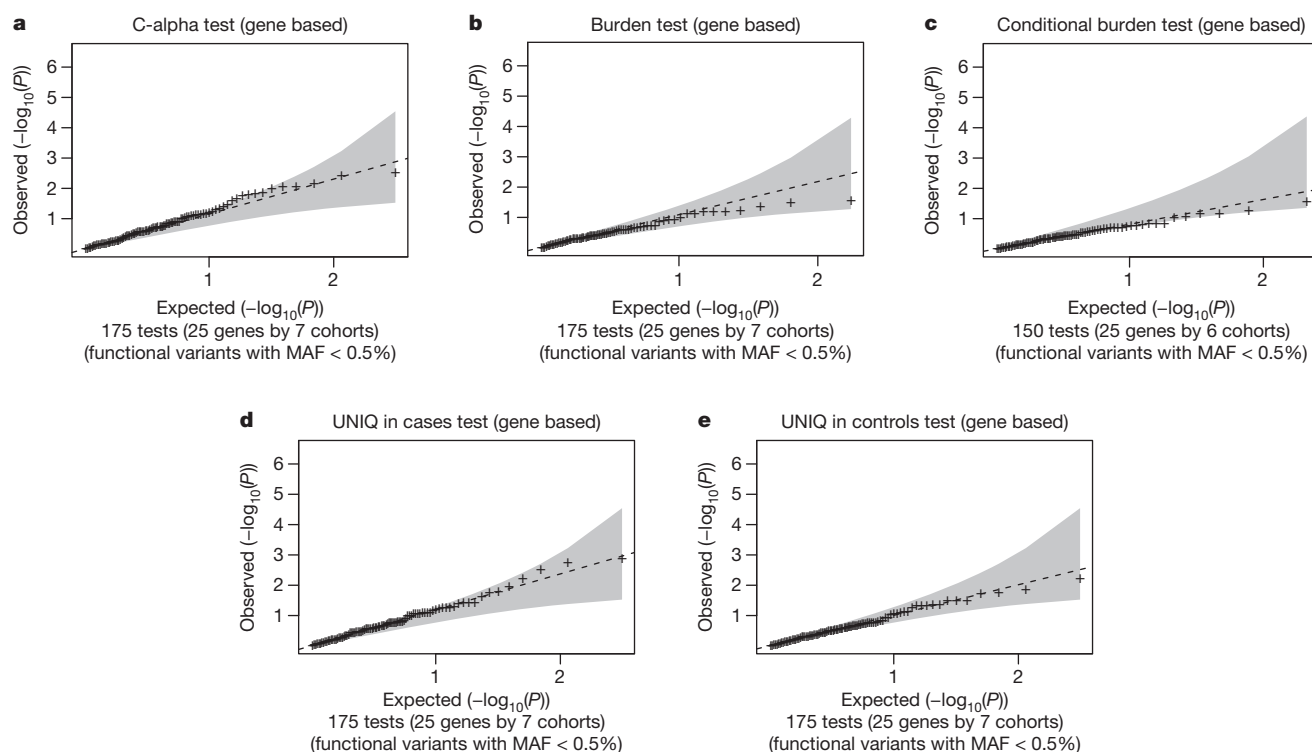


Figure 1 | Association analyses of discovered rare functional variants in autoimmune diseases. We define rare functional variants as MAF < 0.5% in 17,019 controls and predicted nonsynonymous, premature-stop or splice-site annotation. Quantile–quantile plots compare observed versus expected test-statistic distributions, with shading indicating 99% confidence intervals. Full results are available in Supplementary Data. Each of six individual diseases, and all autoimmune diseases combined, were tested as phenotypes. **a**, Gene-based C-alpha test (25 genes by 7 phenotypes, $n = 41,911$ subjects) allowing for both risk and protective effects for rare functional variants. Singleton variants pooled into a single binomial count per phenotype. **b**, Gene-based burden tests (25 genes by 7 phenotypes, $n = 41,911$ subjects) comparing summed allele counts for rare functional variants in cases versus controls with Fisher's exact test.

c, Conditional gene-based burden test (25 genes by 6 phenotypes, $n = 32,806$ subjects): rare functional-variant allele counts are summed for each individual per gene and introduced in a logistic regression, including ImmunoChip covariates for multiple independent top (common) variant signals selected on the basis of a stepwise regression (down to $P > 10^{-4}$). The psoriasis phenotype was not tested as most samples do not have ImmunoChip data. **d**, Count of case-unique rare alleles (UNIQ) tests (25 genes by 7 phenotypes, $n = 41,911$ subjects): compares the number of rare functional variants only observed in cases with the distribution of this value upon random permutation (10,000 times) of the phenotypes. **e**, Count of control-unique rare alleles (UNIQ) tests: same as **d** but for rare functional variants uniquely observed in controls.

assays; (5) proportions of rare, and of known, variants were similar to those found by other large sequencing studies, and we identified no common or low-frequency novel variant sites.

We first attempted to identify any low-frequency or rare variants of larger effect. We performed for each coding-region variant and each of seven phenotypes (including all autoimmune disease cases combined) a single-variant association analysis. Only previously reported loci were observed with common variants ($MAF > 5\%$), as expected. We identified three low-frequency ($MAF = 0.5\text{--}5\%$) and rare (MAF in 17,019 controls $= <0.5\%$) exonic variants with single SNP association $P < 10^{-4}$ (chosen as a partial Bonferroni multiple testing correction for 25 genes and 7 phenotypes, but not correcting for all variants per gene) (Supplementary Table 4 and Supplementary Data). We next analysed low-frequency and rare exonic variants, conditioning on common-variant non-coding signals at each locus, and observed no additional association signals (Supplementary Data). An association between type 1 diabetes and the low-frequency *UBASH3A* SNP rs17114930 was observed, but conditional regression analysis showed this signal to be secondary to a stronger common-frequency variant/haplotype previously identified by GWAS²⁰. We identified novel low-frequency (nearly 'common' as MAF in 17,019 controls $= 4.97\%$) *NCF2* coding-region variant associations with coeliac disease at two SNPs (rs17849502, nonsynonymous; rs17849501, synonymous; in almost complete linkage disequilibrium $r^2 = 0.992$). Both variants were present on the Illumina ImmunoChip, but just failed quality-control criteria in our previous coeliac disease study owing to missing data¹⁹. We replicated the UK findings in 4,313 coeliac cases and 3,954 controls (European samples, Methods; rs17849502 $P = 4.46 \times 10^{-5}$ (Cochran–Mantel–Haenszel test), odds ratio 1.35 (95% CI = 1.17–1.55)). Logistic regression analysis conditioning on rs17849502 in the UK re-sequencing data set revealed no further single-variant coeliac disease association signals below $P < 10^{-4}$. *NCF2* is a component of the neutrophil NADPH oxidase respiratory burst complex. Different disease-causing mutations cause the recessive Mendelian phenotype chronic granulomatous disease. The rs17849502/H389Q variant is also associated with the autoimmune disease systemic lupus erythematosus²¹. Functional studies have shown that the minor allele of rs17849502/H389Q reduces the binding efficiency of *NCF2* to the guanine nucleotide-exchange factor VAV1 (ref. 21). These data now implicate a disease mechanism of impaired neutrophil function in coeliac disease, a condition previously thought to be of predominantly B- and T-cell-mediated immunopathogenesis, and where neutrophils may have a role in regulating adaptive immunity²².

We noted that even with $\sim 7,000$ cases and $\sim 17,000$ controls the power to detect association signals using single-variant tests for variants ($MAF < 0.5\%$) of modest effect (for example, odds ratio < 3) is limited (Supplementary Fig. 2) and therefore we performed gene-based pooled-variant association tests to better detect the combined effect of multiple variants. We defined coding-region variants as functional candidates if the variants were rare (MAF in 17,019 controls $= <0.5\%$) and predicted to be of potential functional impact (nonsynonymous, premature stop, splice-site altering; see Methods). We pooled variants (by gene) in analyses to detect different scenarios (Fig. 1 and Supplementary Data), including the C-alpha test, which can detect a combination of risk and protective variants; burden tests to detect either an excess of risk variants in cases or protective variants in controls; a modified version of the burden test using conditional regression and common-variant non-coding signals at a locus as covariates; a test to detect an excess of rare variants seen uniquely in cases (the case or control unique tests being particularly suitable for the study of the large numbers of singleton and doubleton variants we observe); and a test to detect an excess of rare variants seen uniquely in controls. The distribution of association statistics for all five pooled gene tests across each of the six or seven phenotypes tested was consistent with the global null of no association.

On the basis of these results, in the largest (to the best of our knowledge) human disease sample sequencing study to date, we find little

support for a significant impact of rare coding-region variants in known risk genes for the autoimmune disease phenotypes tested. Our data provide little stimulus in support of large-scale whole-exome sequencing projects in common autoimmune diseases. Using average genetic-effect estimates from our data (Methods), over all loci and phenotypes we have tested, we estimate that rare variants contribute to less than 3% of the heritability explained by common variants at these known risk loci²³.

METHODS SUMMARY

Sequencing. DNA (corresponding to exonic sequence of 25 autoimmune disease risk genes) was PCR-amplified in a multiplexed microfluidics assay (Fluidigm Access Array). PCR amplicons from a sample were pooled, and barcoded with one of 1,536 unique ten-base-pair sequences. Libraries of 1,536 samples were sequenced on Illumina HiSeq instruments. Reads were aligned to the GRCh37 human reference and SNVs and small indels called. Samples and called variants were extensively filtered on the basis of call rate and other criteria. Selected variants were validated by Sanger dideoxy sequencing. Genotype data from Illumina ImmunoChip array-based genotyping was merged with Fluidigm sequencing-based genotypes.

Statistical analysis. Statistical analysis was performed in R, and using PLINK/SEQ software.

Full Methods and any associated references are available in the online version of the paper.

Received 27 February; accepted 8 April 2013.

Published online 22 May 2013.

- Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Gibson, G. Rare and common variants: twenty arguments. *Nature Rev. Genet.* **13**, 135–145 (2012).
- Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
- Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genet.* **43**, 1066–1073 (2011).
- Momozawa, Y. *et al.* Resequencing of positional candidates identifies low frequency *IL23R* coding variants protecting against inflammatory bowel disease. *Nature Genet.* **43**, 43–47 (2011).
- Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
- Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T. L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234–237 (2013).
- Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genet.* **44**, 483–489 (2012).
- Park, J. H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genet.* **42**, 570–575 (2010).
- Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42**, 565–569 (2010).
- Coventry, A. *et al.* Deep resequencing reveals excess rare variants consistent with explosive population growth. *Nature Commun.* **1**, 131 (2010).
- Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104 (2012).
- Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
- Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
- Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nature Genet.* **42**, 985–990 (2010).
- Jordan, C. T. *et al.* Rare and common variants in *CARD14*, encoding an epidermal regulator of NF-kappaB, in psoriasis. *Am. J. Hum. Genet.* **90**, 796–808 (2012).
- Diogo, D. *et al.* Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. *Am. J. Hum. Genet.* **92**, 15–27 (2013).
- Lesage, S. *et al.* *CARD15/NOD2* mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am. J. Hum. Genet.* **70**, 845–857 (2002).
- Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genet.* **43**, 1193–1201 (2011).
- Barrett, J. C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature Genet.* **41**, 703–707 (2009).
- Jacob, C. O. *et al.* Lupus-associated causal mutation in neutrophil cytosolic factor 2 (*NCF2*) brings unique insights to the structure and function of NADPH oxidase. *Proc. Natl Acad. Sci. USA* **109**, E59–E67 (2012).
- Kolaczowska, E. & Kubas, P. Neutrophil recruitment and function in health and inflammation. *Nature Rev. Immunol.* **13**, 159–175 (2013).

23. Liu, D.J. & Leal, S.M. Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. *Am. J. Hum. Genet.* **91**, 585–596 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The study was primarily funded by the Medical Research Council (MRC G1001158 to D.A.v.H. and V.P.), with further funding from Coeliac UK (to D.A.v.H.). We thank C. Wijmenga and G. Trynka for sharing ImmunoChip data, and the International Multiple Sclerosis Genomics Consortium for ImmunoChip data and samples. J.N.B. and R.C.T. are supported by MRC grant G0601387. This research was supported by the National Institutes for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The study was supported by the Cambridge NIHR Biomedical Research Centre. We thank E. Gray and D. Jones (Wellcome Trust Sanger Institute) for sample preparation. We acknowledge use of DNA from The UK Blood Services collection of Common Controls (UKBS-CC collection), funded by the Wellcome Trust grant 076113/C/04/Z and by NIHR programme grant to NHS Blood and Transplant (RP-PG-0310-1002). The collection was established as part of the Wellcome Trust Case Control Consortium (WTCCC). We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the UK MRC grant G0000934 and the Wellcome Trust grant 068545/Z/02. We thank nurses and doctors for recruiting autoimmune thyroid disease (AITD) subjects into the AITD National Collection, funded by the Wellcome Trust grant 068181. We acknowledge use of DNA from the Cambridge BioResource. We acknowledge use of DNA from the Juvenile

Diabetes Research Foundation (JDRF)/Wellcome Trust Case-Series (GRID), funded by JDRF and the Wellcome Trust (grant references JDRF 4-2001-1008 and WT061858). The subjects were recruited in the UK by D. Dunger and his team with support from the British Society for Paediatric Endocrinology and Diabetes. The samples were prepared and provided by the JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, University of Cambridge, UK. Psoriasis samples used were based on the WTCCC2 GWAS clinical panel, for which we thank D. Burden, C. Griffiths, M. Cork and R. McManus. Finally, we would like to thank all autoimmune disease and control subjects for participating in this study.

Author Contributions D.A.v.H. designed and led the study. K.A.H. coordinated wet laboratory work, with K.A.H., V.M., N.A.B. and E.W. performing DNA sample preparation, Fluidigm PCR amplification, sample barcoding, MiSeq library validation and Sanger sequencing preparation. HiSeq sequencing was performed by M.M. and E.P. D.A.v.H., V.P. and M.S. performed bioinformatics and statistical analyses. All other authors contributed to diverse aspects of sample collection, phenotyping, DNA preparation, ImmunoChip data production or specific analyses. D.A.v.H. and V.P. drafted the manuscript, which all authors reviewed.

Author Information Genome data has been deposited at the European Genome-phenome Archive (<http://www.ebi.ac.uk/ega/>), which is hosted at the EBI, under accession number EGAS00001000476. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A.v.H. (d.vanheel@qmul.ac.uk) or V.P. (v.plagnol@ucl.ac.uk).

METHODS

Gene selection. All genes studied (listed in Supplementary Table 3) were risk loci for at least two phenotypes, had a known immune system function, were from loci with only a single strong candidate immune gene (or all immune genes were selected at four loci: *IL18R1*, *IL18RAP*, *CTLA4*, *CD28*, *ICOS*, *IL2*, *IL21*; *PTPRK*, *THEMIS*), and all genes and loci were densely genotyped with all 1000 Genomes pilot project variants on the Illumina ImmunoChip (for design of this chip, see ref. 19). Additional criteria favouring locus selection were: known multiple independent association signals, risk (not necessarily same variants/haplotype or signal direction) for many autoimmune diseases, fine-mapping or other data strongly suggesting a single candidate gene, and smaller complementary DNA size.

Samples. UK samples for the six component immune disease phenotypes have been described in previous publications (which also contain full details of Ethics Committee approvals)^{19,20,24–27}, as have the three control populations^{19,28}. Informed consent was obtained from all subjects. Individuals with self-reported autoimmune disease were excluded from the UK Blood Services — Common Controls and NIHR Cambridge Biomedical Research Centre Cambridge BioResource controls. Samples with self-stated non-white European ethnicity were excluded (later further confirmed by ImmunoChip-based principal component ethnicity analysis for 32,806 samples). Samples with gross discordance with ImmunoChip genotypes and/or with known gender or genotype-mismatch issues from previous GWAS were excluded. Samples with known duplicates or relatedness (as distant as first cousins) were excluded, relatedness was later confirmed by ImmunoChip genome-wide identity-by-state analysis and by analysis of multiple rare-variant sharing in Fluidigm sequencing data. Additional independent European samples genotyped for rs17849502 (4,313 coeliac cases and 3,954 controls) were previously described¹⁹. **Wet-lab.** PCR primers were designed for all RefSeq exons of 26 genes, and amplicons selected to be 150–200 base pairs (bp) in size. There was minor primer design dropout at *IL18R1*, *STAT4*, *THEMIS* and *ZMIZ1*, although >94% of exon sequence was still covered at these genes. Variant calls at the gene *YDJC* later proved unreliable with highly biased allele depths at heterozygote sites, probably due to the very high exon GC content (~70%), and this gene was not further analysed nor is it discussed elsewhere in this study. The total length of (overlapping) amplicons was 95,927 bp; with primers removed (still overlapping) 72,612 bp; and with primers removed and unique sequence 58,550 bp. PCR amplification was performed using 50 ng genomic DNA per sample on the 48 sample/plate Fluidigm microfluidic Access Array system. PCR primers for 511 PCR reactions were pooled up to 12-plex per well in 48 pools. Individual per sample per pool PCR reactions took place in ~35-nl reaction chambers with ~300 DNA haplotypes per reaction. All pools per sample were combined. Each sample's pool was then individually barcoded in a second PCR reaction with one of 1,536 10-bp Fluidigm-designed unique barcodes (Fluidigm unidirectional sequencing protocol).

Sequencing. Thirty-four libraries (each of 1,536 barcoded samples) were generated. Libraries were first sequenced on an Illumina MiSeq for rapid quality control of the barcoding step, and to optimize loading concentrations/cluster density. Libraries were then sequenced one per lane using 101-bp paired-end reads and an 11-bp index read (the last base of each read being only used for chemistry cycle phasing purposes) on Illumina HiSeq sequencers. Lanes were repeated if target cluster density or target clusters passing filter were not achieved. Individual samples were de-multiplexed by Illumina CASAVA software, allowing zero mismatches per 10-bp barcode. Sanger sequencing was performed on PCR products using an ABI 3730xl DNA analyser and ABI big dye terminator 3.1 cycle chemistry. We sequenced all samples with rare-variant allele genotypes, and a control sample, for the 74 sites selected.

Bioinformatics. PCR primers were trimmed from the 5' end of individual reads using a modified version of btrim²⁹. Trimmed sequences were aligned to the GRCh37 human reference genome using gapped quality-aware alignment, and base call quality recalibration implemented in Novoalign V2.07.18 with settings '-t 100 -H -g 65 -x 7 -o FullNW'. Data were realigned against known (1000 Genomes and Mills-Devine 2-hit) indels and per-sample called indels. SNPs were called using GATK 1.6-5 and settings '-min_base_quality_score 15 -stand_call_conf 30 -baq CALCULATE_AS_NECESSARY -glm SNP -baqGapOpenPenalty 65 -downsampling_type BY_SAMPLE -downsample_to_coverage 250' and then hard filtered using GATK settings 'QUAL<80.0 DP<20 MQ<40.0 QD<2.0 MQRankSum<-12.5 HRun>5' (several other recommended best practice GATK settings were not appropriate for PCR amplicon data), and around indels. Small indels (up to 15-bp gaps from Novoalign) were called using GATK and settings '-min_base_quality_score 15 -stand_call_conf 30 -baq CALCULATE_AS_NECESSARY -glm INDEL -baqGapOpenPenalty 65 -downsampling_type BY_SAMPLE -downsample_to_coverage 250' and then hard filtered using GATK settings 'QUAL<80.0 DP<20 QD<2.0' (several other recommended best-practice GATK settings were not appropriate for PCR amplicon data). The most important of these settings were likely to be calling genotypes as missing with

sequencing depth <20 high-quality bases and the minimum Phred 15 recalibrated base call quality score to define high-quality bases. Both SAMtools and VCFtools software were also used to process data. SNP genotypes (including non-reference genotypes) were called at all 58,550 bases of amplicon sequence. Samples with <57,600 SNP genotype calls (98.4%, a threshold determined by inspection of the call rate plot) were removed and scheduled for repeat processing. Clusters of very close non-reference genotypes in an individual sample were removed. Non-reference genotype sites were then identified across all samples, and VCF-level data reduced to variants at polymorphic sites (in one or more samples). A combined VCF file of all polymorphic sites and samples was then loaded into PLINK/SEQ v0.09. Multiple-step filtering based on call rate per sample and call rate per variant site was applied, with final requirements >99.95% call rate per sample and per variant site. Lower call rate samples at this stage were also scheduled for repeat processing. We removed variants if the sum of heterozygote genotype allele depths was <25% or >75%. The final filtered data was then exported to a VCF file containing all variants and samples for analysis in R. ImmunoChip data was loaded into Illumina GenomeStudio software from .idat files, and all samples called together in GenomeStudio using the cluster settings as previously described¹⁹. Data were merged with HapMap Phase 3 genotypes, principal component analysis performed, and the first two principal components used to validate ethnicity (Supplementary Fig. 3).

Barcode and sequencing amplicon performance. Barcode evenness was excellent, with typically 99.0% of the 1,536 barcodes producing pass-filter read numbers that were between 0.033% and 0.13% of the total pass-filter reads per lane (0.065% expected), with most of the failing barcodes tagging known water-negative control samples or (based on repeat amplification with a different barcode) due to poor DNA quality. Amplicon evenness was good, and for many genotype calls we were required to downsample data to 250 bases per site per sample (Supplementary Fig. 1). However, 10 of 511 amplicons effectively failed PCR. In a typical analysis of 100 high-quality samples, 2% of the 58,550 unique amplicon bases had a minimum mean read-depth of <20, nearly all accounted for by the 10 failing amplicons.

Variant annotation. Annotation of all variants was first performed using ANNOVAR (Feb 2013) and the GENCODE V14 data set. Coding variants were identified. Rare functional variants were identified based on stop, frameshift indel, nonsynonymous (SNV or 3n indel) or splice predictions. We performed an additional layer of annotation for high confidence loss of function mutations, using the methods described in ref. 30. The Variant Effect Predictor (VEP v2.5) tool from Ensembl was modified to produce custom annotation tags and additional loss of function (LOF) annotations. The additional LOF annotation was applied to variants which were annotated as STOP_GAINED, SPLICE_DONOR_VARIANT, SPLICE_ACCEPTOR_VARIANT, and FRAME_SHIFT and flagged if any filters failed. Filters included: LOF is the ancestral allele; exon is surrounded by non-canonical splice site (that is not AG/GT); LOF removes less than 5% of remaining protein; LOF is rescued by nearby start codon which results in less than 5% of protein truncated; transcript only has one coding exon; splice-site mutation within intron smaller than 15 bp; splice site is non-canonical OR other splice site within same intron is non-canonical; unable to determine exon/intron boundaries surrounding variant. A LOF variant is predicted as high confidence if there is one transcript that passes all filters, otherwise it is predicted as low confidence. We noted that LOF mutations were seen in 21 out of 25 genes, all were heterozygous genotypes, and mainly (87 out of 97) as singletons or doubletons in the 41,911 samples (Supplementary Table 3).

Statistical analysis. Most analysis was performed in R using custom code (available on request). For tests using permutations (C-alpha, UNIQ-cases and UNIQ-controls in Fig. 1), we randomly permuted in R the case-control status 10,000 times. The unconditional burden test (Fig. 1b) used a Fisher's exact test. Conditional burden tests used the glm function in R, including selected ImmunoChip common variants as covariates (selection based on a stepwise regression analysis up to 10⁻⁴). For the C-alpha statistic computation (Fig. 1a), the expected proportion of rare alleles in the case-control cohorts was set to the proportion of cases and controls. Figure 1 was generated using the fact that under the null of no association -2log(P) is distributed as chi-squared with 2 degrees of freedom. PLINK/SEQ v0.09 (<http://atgu.mgh.harvard.edu/plinkseq/index.shtml>) was used for Ti/Tv statistics, and to confirm findings of R analyses (not shown). We used PLINK/SEQ for the genotype concordance analysis between ImmunoChip and Fluidigm-sequencing data. Discordant calls were observed at 169 of 2,985,255 (0.0056%) genotypes, occurring at 36 out of 91 polymorphic variant sites present in both data sets. We inspected Illumina ImmunoChip R theta intensity plots for the discordant genotypes, and observed 8 discordant genotypes to be likely due to ImmunoChip data mis-clustering, and 11 discordant genotypes to be due to a third or fourth observed allele in the high-throughput sequencing data. At the sites with third and fourth alleles, we note the ImmunoChip array assays can

only call two alleles, therefore is not possible to determine whether these sequence genotype calls are real or errors. R code used for analysis is available from V.P.

Estimation of average genetic effect contributed by rare variants. For each combination of locus by disease, we combined all rare functional variants (frequency < 0.5% in 1,000 Genomes/NHLBI data sets and nonsynonymous, LOF or splicing) in a burden statistic X and computed the combined frequency of X in the sample. Using a logistic regression model with the disease phenotype as outcome, we estimated the odds ratio associated with the burden variable X. This knowledge of frequency and odds ratio for the burden variable X enables the estimation of the average genetic effect (AGE, as defined in ref. 23) version of the variance explained. We then compared this variance at each combination of locus/gene with the variance explained by what we consider to be a typical common variant association (odds ratio 1.2, MAF 20%, assuming a single common variant per locus). To deal with the uncertainty in estimated odds ratio and obtain a confidence interval for this value, we randomly sampled the odds ratio from their estimated distribution for each pair of disease/locus. Averaging over the 150 combinations of 6 diseases by 25 loci, we estimate the ratio of heritability explained for all rare variants by all common variants to have a mean value of 1.6%, with a confidence interval of (1.2–2.3%). It is pointed out in ref. 23 that the AGE estimate can underestimate the true explained variance by rare variants. Nevertheless, assuming that rare variants are generally all risk or all protective at a given gene, their simulations also show that the underestimation is limited, in the range of a 25% decrease. Taking this conservative estimate of the under-estimation level, we find the upper bound of the 95% of the confidence interval to be 3.05%. Hence, our data indicate that the aggregate contribution of rare

variants to the heritability (<0.5% MAF, and averaged over these loci/diseases) is unlikely to exceed approximately 3% of the heritability assigned to common variants. We acknowledge that a much larger underestimation (and therefore a much larger heritability explained for rare variants) is possible in the presence of a combination of high risk and highly protective rare variants at the same locus. Although we cannot exclude such scenario, it is unlikely to be widespread. We also assumed in our estimates that rare variants act additively at the log scale. Although this assumption is standard, we cannot exclude that a combination of rare variants results in a much stronger predictive outcome than rare variants individually, hence underestimating the heritability associated with rare variants.

24. Cooper, J. D. *et al.* Seven newly identified loci for autoimmune thyroid disease. *Hum Mol Genet* **21**, 5202–5208 (2012).
25. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
26. Sawcer, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
27. Tsoi, L. C. *et al.* Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet* **44**, 1341–1348 (2012).
28. Dendrou, C. A. *et al.* Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. *Nat Genet* **41**, 1011–1015 (2009).
29. Kong, Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* **98**, 152–153 (2011).
30. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).

Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells

Alex K. Shalek^{1*}, Rahul Satija^{2*}, Xian Adiconis², Rona S. Gertner¹, Jellert T. Gaublomme¹, Raktima Raychowdhury², Schraga Schwartz², Nir Yosef², Christine Malboeuf², Diana Lu², John J. Trombetta², Dave Gennert², Andreas Gnirke², Alon Goren^{2,3}, Nir Hacohen^{2,4}, Joshua Z. Levin², Hongkun Park^{1,2} & Aviv Regev^{2,5}

Recent molecular studies have shown that, even when derived from a seemingly homogenous population, individual cells can exhibit substantial differences in gene expression, protein levels and phenotypic output^{1–5}, with important functional consequences^{4,5}. Existing studies of cellular heterogeneity, however, have typically measured only a few pre-selected RNAs^{1,2} or proteins^{5,6} simultaneously, because genomic profiling methods³ could not be applied to single cells until very recently^{7–10}. Here we use single-cell RNA sequencing to investigate heterogeneity in the response of mouse bone-marrow-derived dendritic cells (BMDCs) to lipopolysaccharide. We find extensive, and previously unobserved, bimodal variation in messenger RNA abundance and splicing patterns, which we validate by RNA-fluorescence *in situ* hybridization for select transcripts. In particular, hundreds of key immune genes are bimodally expressed across cells, surprisingly even for genes that are very highly expressed at the population average. Moreover, splicing patterns demonstrate previously unobserved levels of heterogeneity between cells. Some of the observed bimodality can be attributed to closely related, yet distinct, known maturity states of BMDCs; other portions reflect differences in the usage of key regulatory circuits. For example, we identify a module of 137 highly variable, yet co-regulated, antiviral response genes. Using cells from knockout mice, we show that variability in this module may be propagated through an interferon feedback circuit, involving the transcriptional regulators Stat2 and Irf7. Our study demonstrates the power and promise of single-cell genomics in uncovering functional diversity between cells and in deciphering cell states and circuits.

To characterize the extent of expression variability on a genomic scale and decipher its functional implications, we used single-cell RNA sequencing (RNA-Seq) to profile a temporal snapshot of the BMDC response to lipopolysaccharide (LPS). This is an attractive model system for single-cell analyses for several reasons. First, LPS, a component of Gram-negative bacteria and a ligand of Toll-like receptor 4, strongly synchronizes cellular responses and mitigates temporal phasing¹¹. Second, LPS activation evokes a robust transcriptional program that has been extensively investigated at the population level¹². Third, LPS stimulation should increase the correlation between mRNA and protein levels for induced genes, thus reducing a potentially confounding factor¹³. Finally, differentiated BMDCs are post-mitotic, largely removing cell cycle-dependent transcriptional variation³.

We stimulated BMDCs with LPS and collected single cells after four hours¹² (Supplementary Information). Using SMART-Seq⁹, we constructed complementary DNA libraries from 18 single BMDCs (S1–S18), three replicate populations of 10,000 cells, and two negative controls (empty wells), and sequenced each to an average depth of 27 million read pairs. Negative control libraries failed to align (<0.25%) to the

mouse genome, and were discarded from all further analyses. Library quality metrics, such as genomic alignment rates, ribosomal RNA contamination, and 3' or 5' coverage bias, were similar across all libraries (Supplementary Table 1). We estimated expression levels for all University of California Santa Cruz (UCSC)-annotated genes using RSEM (RNA-Seq by expectation maximization)¹⁴ (Supplementary Table 2) and discarded genes that were not appreciably expressed (transcripts per million (TPM) > 1) in at least three individual cells, retaining 6,313 genes for further analysis.

Although the gene expression levels of population replicates were tightly correlated with one another (Pearson $r > 0.98$, log-scale; Fig. 1a), there were substantial differences in expression between individual cells ($0.29 < r < 0.62$, mean: 0.48; Fig. 1b and Supplementary

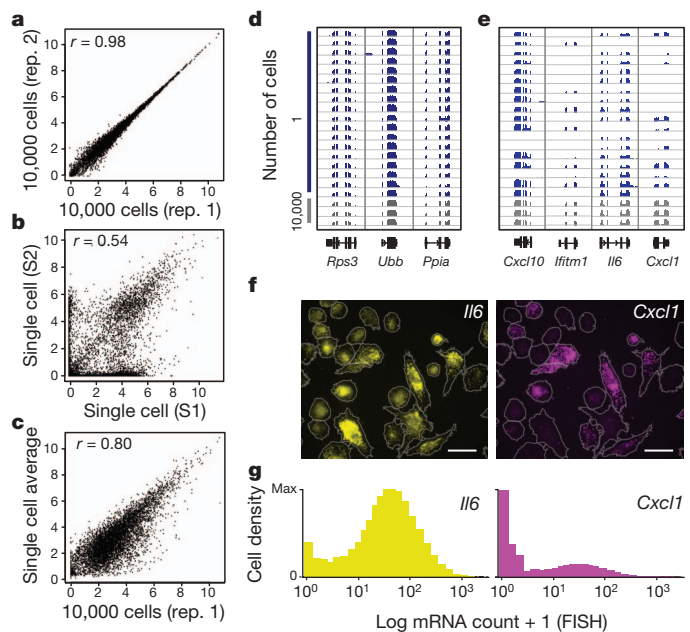


Figure 1 | Single-cell RNA-Seq of LPS-stimulated BMDCs reveals extensive transcriptome heterogeneity. a–c, Correlations of transcript expression levels (x and y-axes: log-scale TPM + 1) between two 10,000-cell population replicates (rep. 1 and rep. 2) (a), two single cells (S1 and S2) (b), and the ‘average’ single cell and a population (c). d, e, RNA-Seq read densities in single cells (blue) and population replicates (grey) for three non-variable genes (d) and four variable ones (e). f, g, RNA-FISH of representative transcripts. Optical micrographs (cell boundaries; grey outlines) and maximum-normalized distributions of expression levels from a RNA-FISH co-staining ($n = 3,193$ cells) for *Il6* (yellow) and *Cxcl1* (magenta). Scale bars, 25 μ m.

¹Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. ²Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ³Department of Pathology & Center for Systems Biology and Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. ⁴Center for Immunology and Inflammatory Diseases & Department of Medicine, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. ⁵Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02140, USA.

*These authors contributed equally to this work.

Fig. 1). Despite this extensive cell-to-cell variation, expression levels for an 'average' single cell correlated well with the population samples ($0.79 < r < 0.81$; Fig. 1c and Supplementary Fig. 1).

We used RNA-fluorescence *in situ* hybridization (RNA-FISH), an amplification-free imaging technique², to verify that heterogeneity in our single-cell expression data reflected true biological differences, rather than technical noise associated with the amplification of small amounts of cellular RNA. For 25 genes, selected to cover a wide range of expression levels, the variation in gene expression detected by RNA-FISH closely mirrored the heterogeneity observed in our sequencing data (Fig. 1d–g and Supplementary Fig. 2). For example, expression of housekeeping genes (such as β -actin (*Actb*) and β 2-microglobulin (*B2m*)) matched a log-normal distribution in both single-cell RNA-Seq and RNA-FISH measurements, consistent with previous studies¹. By contrast, many genes involved in the LPS response, although highly expressed on average, showed significantly greater levels of heterogeneity, with expression levels deviating $\sim 1,000$ -fold between individual cells in extreme cases (Fig. 1e–g).

More generally, we observed that single-cell variability existed across a wide range of population expression levels (Fig. 2a). Of the 522 most highly expressed genes (single-cell average TPM > 250; Fig. 2a, unshaded region, and Supplementary Table 3), 281 had low cell-to-cell variability (coefficient of variation (CV , σ/μ) < 0.25; Supplementary Information) and were well described by log-normal distributions (RNA-Seq: Fig. 2b, c, top, RNA-FISH (*Actb*, *B2m*): Supplementary Fig. 2). These 281 genes were enriched for housekeeping genes, encoding ribosomal and other structural proteins (Supplementary Tables 2 and 3; Bonferroni-corrected $P = 1.5 \times 10^{-6}$), consistent with previous findings in yeast¹⁵ and mammalian cells¹.

Notably, however, 185 of the remaining 241 ($CV > 0.25$; Supplementary Information) highly expressed genes had bimodal expression patterns (Fig. 2b, c, bottom): mRNA levels for these genes were high in many of the cells, but were at least an order of magnitude lower (often very low or undetectable) than the single-cell average in three or more cells. We independently verified this disparity by RNA-FISH (for example, *Cxcl1*, *Cxcl10* and *Ifit1*; Fig. 1f, g and Supplementary Fig. 2), confirming that it was not a result of technical noise. This variable set included both antiviral and inflammatory response genes, and was highly enriched for genes in which expression was increased by at least twofold after LPS stimulation at the population level¹⁶ ($P = 2.7 \times 10^{-7}$; hypergeometric test; Supplementary Table 2). Still, bimodal expression was not a universal feature of immune response transcripts; some key chemokines and chemokine receptors (*Ccl3*, *Ccl4* and

Ccl2), cytokines (*Cxcl2*), and signalling molecules (*Tank*) were highly expressed in every cell (Supplementary Fig. 3), indicating that all cells were indeed activated by LPS.

This degree of variation in expression for highly expressed (on average) transcripts has not been observed in previous reports^{7–10}. For example, examination of published single-cell RNA-Seq data sets of human embryonic stem cells⁹ (Fig. 2a), mouse embryonic stem cells, and terminally differentiated fibroblasts¹⁰ (Supplementary Fig. 4) revealed far less heterogeneity in expression for highly abundant (population average) genes. Similarly, studies of protein expression in mid-log yeast cells and dividing human cell lines^{15,17} did not find such bimodality in (on average) highly expressed genes. We thus proposed that widespread variability in single-cell gene expression may reflect functionally important differences in the stimulated BMDC population.

Furthermore, we found that splicing patterns also showed previously unobserved levels of heterogeneity across single cells. Specifically, for genes that have multiple splice isoforms at the population level, individual cells predominantly expressed one particular isoform. We calculated the frequency (percentage spliced in (PSI)) of previously annotated splicing events in each of our samples using MISO¹⁸, a Bayesian framework for calculating isoform ratios (Supplementary Table 4). Although the population-derived estimates were highly reproducible, single cells exhibited significant variability in their exon-inclusion frequencies (Fig. 3a, b).

We considered the possibility that PCR amplification (intrinsic to the library preparation process) could potentially produce an over-estimation of isoform regulation variability, particularly for weakly expressed transcripts¹⁹. However, even when we limited our analysis to 89 alternatively spliced exons ($0.2 < \text{population PSI} < 0.8$) that were very highly expressed within a single cell (single cell TPM > 250; Supplementary Information), we still observed the same variability in splicing patterns among individual cells, with highly skewed expression towards a single splice variant (Fig. 3b). We obtained similar results when we generated three additional single-cell cDNA libraries using a slightly modified SMART-Seq protocol (Supplementary Information) in which a four-nucleotide barcode was introduced onto each RNA molecule during reverse transcription¹⁹, enabling us to estimate the number of unique RNA transcripts that existed before PCR (Supplementary Figs 5 and 6 and Supplementary Information).

To the best of our knowledge, single-cell variation in splicing patterns has rarely been studied for individual genes, and never been analysed on a genomic scale. One recent report²⁰ used RNA-FISH to

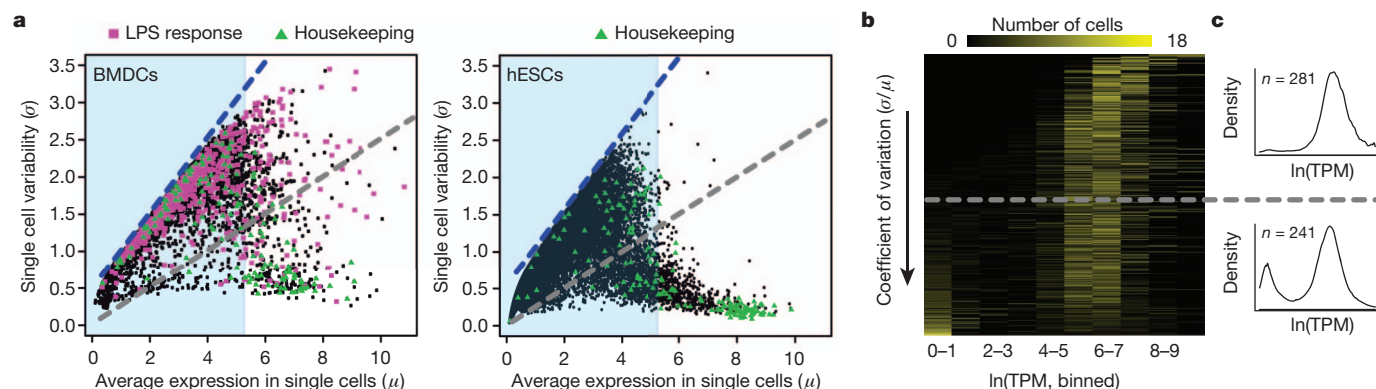


Figure 2 | Bimodal variation in expression levels across single cells.

a, Relationship between average expression level in single cells (μ , x axis) and standard deviation (σ , y axis) for 6,313 genes (Supplementary Table 2). Blue dashed line denotes maximum theoretical σ for an average expression level (Supplementary Information). Grey dashed line denotes the constant coefficient of variation (CV , $\sigma/\mu = 0.25$). Magenta represents immune response genes; green denotes housekeeping genes; light blue shaded region represents single-cell average TPM < 250. **b**, Cellular heterogeneity for the 522 most

highly expressed genes (single-cell average; Supplementary Table 3). Each row represents a discretized histogram for a single gene (sorted by CV from low to high (top to bottom)). Colour represents the number of cells (yellow: 18 cells; black: 0) that express the gene at the noted level. Grey dashed line denotes the constant CV (0.25) highlighted in (a). **c**, Averaged expression density distributions for the 281 low-variability genes (top) and the 241 highly variable genes (bottom).

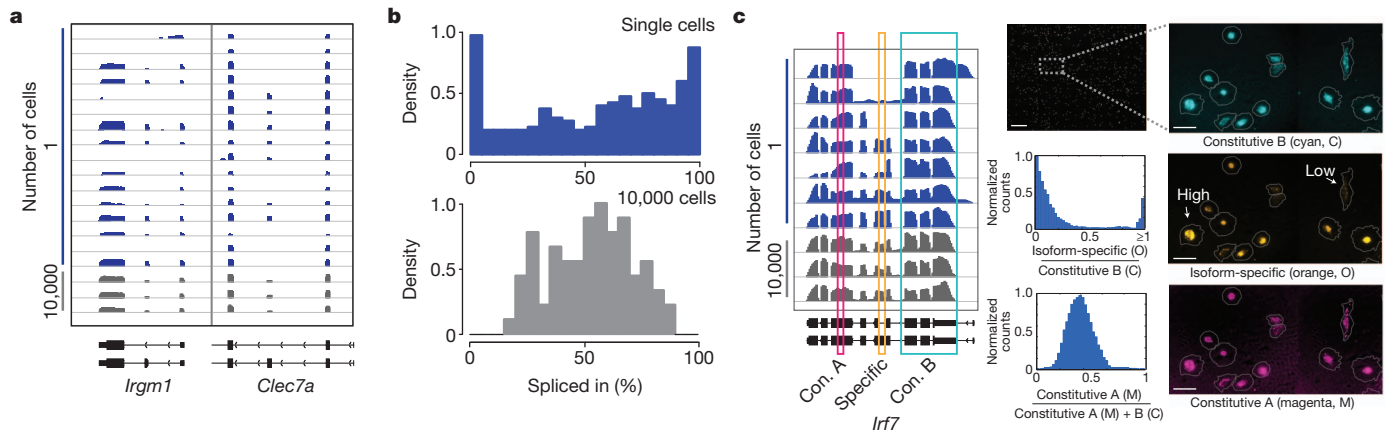


Figure 3 | Variation in isoform usage between single cells. **a**, RNA-Seq read densities in single cells (blue) and population replicates (grey) for two illustrative loci, each with two different isoforms (bottom). **b**, Distributions of exon inclusion (PSI scores, x axis) for alternatively spliced exons of highly expressed genes (single-cell TPM > 250) in individual cells (blue histogram, top) and populations (grey histogram, bottom). **c**, Left, RNA-Seq read densities for *Ir7* (only cells in which the transcript is expressed are shown). Coloured boxes mark exons analysed by RNA-FISH. Right, RNA-FISH images from

simultaneous hybridization with probes for two constitutive (con.) regions of the transcript (A: cyan (C); B: magenta (M)) and one alternatively spliced exon (specific: orange (O)). White arrows (middle) highlight two cells with high levels of *Ir7*, but opposite preferences for the alternatively spliced exon. Histograms show global abundance ratios for isoform-specific and constitutive probes (cells with less than five constitutive counts have been excluded; $n = 490$ cells; bottom histogram deviates from 0.5 owing to probe design; see Supplementary Information). Scale bars, 250 μ m (left); 25 μ m (right).

study variation in alternative isoform usage in two genes, and observed lower levels of isoform variability across single cells (the levels of heterogeneity differed in different cell types). Another study that used fluorescent reporters to quantify single-cell exon inclusion levels for one gene discovered highly variable and bimodal splicing patterns²¹.

To independently verify the existence of extensive differences in isoform ratios between cells, we designed RNA-FISH probes targeting constitutive and isoform-specific exons in two genes²⁰ (*Ir7* and *Acpp*; Fig. 3c and Supplementary Figs 7 and 8). We found substantial expression variability in overall *Ir7* levels between individual cells (as reflected by the 'constitutive' probes; Fig. 3c, top and bottom), mirroring our single-cell sequencing results (and further explored below). Furthermore, within each *Ir7*-expressing cell, we observed a bias towards either the inclusion or exclusion of the cassette exon (Fig. 3c and Supplementary Fig. 7, middle; for example, compare 'high' and 'low' marked cells). We obtained comparable results for *Acpp* using two probes designed to detect mutually exclusive alternative final exons (Supplementary Fig. 8).

We next explored the sources and functional implications of expression variability. Bimodality among highly expressed immune

response genes may reflect the presence of distinct cellular subtypes or stochastic differences in the activation of regulatory circuits¹¹. We performed a principal components analysis (Fig. 4a) on our single-cell expression profiles, focusing on the 632 genes that were induced at least twofold in the population-wide response to LPS¹⁶ (Supplementary Table 5). We found two distinct subpopulations, clearly distinguishable by the first principal component (PC1, 15% of the total variation; Fig. 4a). One group of fifteen cells expressed a core set of antiviral and inflammatory defence cytokines (including *Tnf*, *Il1a*, *Il1b* and *Cxcl10*) at extremely high levels (TPM > 1,000), whereas the remaining three cells expressed them at far weaker levels (TPM < 50). Some cell surface proteins (*Ccr7* and *Cd83*) and chemokines (*Ccl22*), which are known markers of BMDC maturation, showed the opposite expression pattern (Fig. 4b and Supplementary Fig. 9).

During maturation, BMDCs switch from antigen-capturing to antigen-presenting cells that prime the adaptive immune system²². Maturation can occur either in response to pathogen-derived ligands (pathogen-dependent maturation), such as LPS, or when clusters of BMDCs are disrupted in culture²² (pathogen-independent maturation).

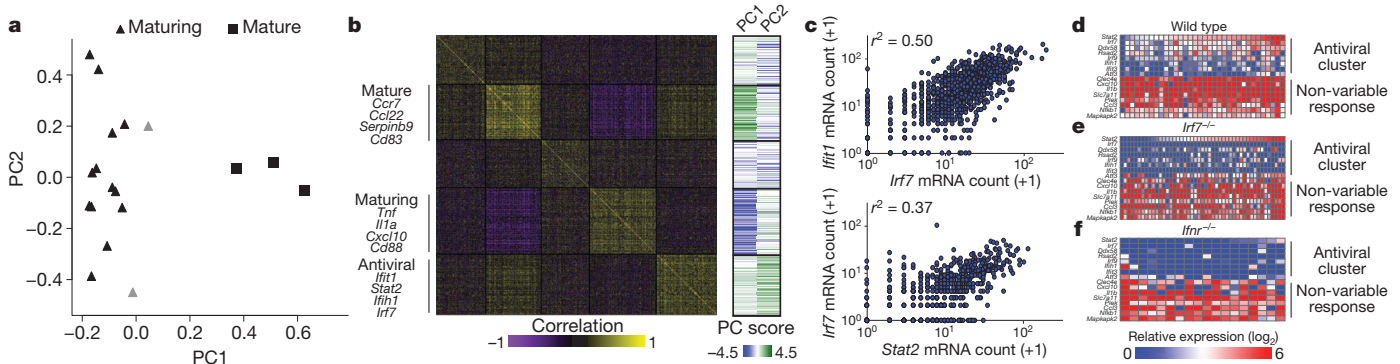


Figure 4 | Analysis of co-variation in single-cell mRNA expression levels reveals distinct maturity states and an antiviral cell circuit. **a**, Principal components analysis of 632 LPS-induced genes. Contributions of each cell (points) to the first two principal components (PC1 and PC2). **b**, Clustered correlation matrix of induced genes. Left, the Pearson correlation coefficients (r) between single-cell expression profiles of every pair of 632 LPS-induced

genes (rows, columns). Right, the projection score (green: high; blue: low) for each gene (row) onto PC1 (left) and PC2 (right). **c**, Confirmation of correlations for *Ir7*-*Stat2* ($n = 655$ cells) and *Ir7*-*Ifit1* ($n = 934$ cells) by RNA-FISH. **d-f**, Expression levels for 16 genes in single BMDCs (columns), measured using single-cell qRT-PCR, in wild type ($n = 36$) (**d**), *Ir7*^{-/-} ($n = 47$) (**e**) and *Ifnr*^{-/-} ($n = 18$) (**f**), at 4 h after LPS stimulation (Supplementary Information).

Both processes lead to induction of maturation markers, but only pathogen-dependent maturation results in co-expression of defence cytokines.

Examining the expression of maturation markers and defence cytokines (Supplementary Fig. 9) suggested that our 18 cells represent two distinct maturity states: (1) 15 cells that were in the early stages of pathogen-dependent maturation (Fig. 4a, 'maturing', triangles; grey triangles, the two cells furthest along in this process); and (2) three cells that probably matured during the culturing process (Fig. 4a, 'mature', squares; pathogen-independent). We further verified the existence of these sub-populations via RNA-FISH (Supplementary Fig. 10), single-cell quantitative reverse transcription PCR (qRT-PCR; Supplementary Fig. 11, Supplementary Information and Supplementary Table 6), and cell sorting based on surface markers identified from the RNA-Seq data (Supplementary Fig. 12 and Supplementary Information). These results highlight that single-cell RNA-Seq can sensitively distinguish between closely related, yet distinct, developmental states, even within the same cell type.

Because differences in cell state explain only a small portion of the observed heterogeneity, we next examined the variation that might arise from the differential activity of regulatory circuits. We reasoned that co-variation across single cells between the mRNA levels of a transcription factor and its targets would represent a potential regulatory interaction, and, furthermore, would suggest that heterogeneity in the regulator's expression may underlie the variability of its targets. Such a correlative approach has successfully identified regulatory connections from population-level transcription profiles measured in different conditions^{2,23}. Here, we attempted to apply it to several single cells in the same condition.

To this end, we calculated the correlation in expression profiles between every pair of induced genes across all single cells, and identified a cluster of 137 genes that varied in a correlated way and were strongly discriminated by the second principal component (PC2, 8% of the variation; Fig. 4a, b). The genes of this cluster included the known antiviral master regulators *Irf7* and *Stat2*, and were highly enriched for members of the antiviral response¹² (60 out of 137 genes, $P = 2.5 \times 10^{-3}$, hypergeometric test; Supplementary Table 5), as well as STAT2 targets¹⁶ (73 out of 137 genes, $P = 4.5 \times 10^{-5}$, hypergeometric test). Most (100 out of 137) of the cluster's genes were bimodally expressed across single cells (Fig. 2c, bottom) despite being strongly expressed at the population level (13 genes TPM > 250; 53 genes TPM > 50). We independently validated a subset of these correlations using single-cell qRT-PCR and RNA-FISH (Fig. 4c, d). Moreover, single-cell qRT-PCR analysis of additional time points demonstrated that these correlations persisted at 6 h as well (Supplementary Discussion and Supplementary Fig. 13).

We hypothesized that bimodal variation in the expression of the cluster's genes may be related to differences in the levels and activities of *Stat2* and *Irf7*. To test this hypothesis, we measured expression of a set of antiviral genes by single-cell qRT-PCR in LPS-stimulated BMDCs from *Irf7* knockout (*Irf7*^{-/-}) mice (Supplementary Information). As expected, this perturbation ablated expression of most of the variable antiviral transcripts in our signature, while leaving non-variable antiviral transcripts relatively unaffected (Fig. 4e). However, *Stat2* expression and variability levels were unaffected by the *Irf7* knockout, indicating that *Stat2* may act either upstream or in parallel to *Irf7* during the response²⁴ (Supplementary Fig. 14). As both *Stat2* and *Irf7* are targets of the interferon-signalling pathway, we stimulated and profiled BMDCs from interferon receptor knockout (*Ifnr*^{-/-}) mice. In these cells, we found markedly reduced expression for both *Stat2* and *Irf7*, as well as all other measured cluster genes (Fig. 4f).

Our analysis provides a proof-of-concept demonstrating how co-variation between transcripts across seemingly homogeneous single cells can help to identify and assemble regulatory circuits. Specifically, in our variable circuit (Supplementary Fig. 14) interferon signalling is required for the induction of *Stat2* and *Irf7*, which, in turn, act to

induce our variable antiviral cluster genes. Our experiments do not definitively determine, however, which component of the circuit causes the observed heterogeneity per se. One compelling possibility is that upstream noise is propagated from the interferon-signalling pathway first to *Stat2* and *Irf7* and then to the target genes^{25,26}. This hypothesis is supported by the variation we observed in *Stat1* and *Stat2* protein levels and nuclear localization (Supplementary Discussion and Supplementary Figs 15 and 16). However, because temporal snapshots of RNA and protein are not always directly comparable (Supplementary Discussion and Supplementary Figs 15 and 16), new strategies for tracing the spatiotemporal dynamics of both proteins and RNA in single living cells are needed to fully test this hypothesis¹¹.

A similar approach could potentially be used to explore the consequences of bimodality in splicing. Even looking at just 18 cells, we witnessed interesting examples of bimodal splicing patterns for genes whose isoforms have distinct functional consequences. For example, the splicing regulators *Srsf3* and *Srsf7* are each known to contain a 'Poison cassette exon' that, when included, targets the RNA for degradation via nonsense-mediated decay²⁷ (Supplementary Fig. 17). Meanwhile, splicing differences in other regulatory genes may further enhance expression diversity: for example, proteins encoded by different isoforms of *Irf7* (Fig. 3c) differentially activate interferon-responsive genes *in vitro*²⁴. These examples suggest that heterogeneity in splicing may represent another layer of response encoding.

In conclusion, our study reveals extensive bimodality in the transcriptional response of BMDCs to LPS, reflected in gene expression, alternative splicing and regulatory circuit activity. Although some variation in expression reflects differences in developmental state, other bimodal patterns reflect the differential activity of an antiviral regulatory circuit in this temporal snapshot. These phenomena allowed us to treat each cell as a 'perturbation system' for reconstructing cell circuits²⁸, even with relatively few cells.

Moreover, our results demonstrate how co-variation across single cells can help dissect and refine gene modules that may be indistinguishable in population-scale measurements. For instance, in a recent population-scale study¹⁶, we identified a large cluster of 808 'late-induced' LPS genes that was enriched for both maturation genes and *Stat*-regulated antiviral genes. These two subsets could not be separated by population-level expression profiles alone¹⁶, but our single-cell data from a single time point clearly distinguishes them. Similarly, the unexpected and prevalent skewing we discovered in alternative splicing between single cells revises our molecular view of this process. Furthermore, although many of our analyses focused on highly expressed genes to reduce the potential influence of amplification noise, our data also revealed substantial bimodality among more moderately expressed transcripts, such as large non-coding RNAs (lincRNAs; Supplementary Fig. 18). This suggests that the low population-level expression of these transcripts²⁹ may sometimes reflect high expression in a small subset of cells as opposed to uniform levels of low expression. Although further technical improvements will be necessary to disentangle these two hypotheses (Supplementary Fig. 5), single-cell measurements should help to facilitate the discovery and annotation of lincRNAs.

Comparing our results to other single-cell RNA-Seq data sets (for example, Fig. 2a and Supplementary Fig. 4) indicates that the source of the analysed tissue (*in vitro* versus *ex vivo*), the biological condition of the individual cells (steady state versus dynamically responding), and the cellular microenvironment all probably influence the extent of single-cell heterogeneity within a system. When applied to complex tissues—such as unsorted bone marrow, developing embryos, tumours and other rare clinical samples—the variability seen through single-cell genomics may help to determine new cell classification schemes, identify transitional states, discover previously unrecognized biological distinctions, and map markers that differentiate them. Fulfilling this potential would require new strategies to address the high levels of noise inherent in single-cell genomics—both technical, owing

to minute amounts of input material, and biological, for example, owing to short bursts of RNA transcription³⁰. Future studies that couple technological advances in experimental preparation with new computational approaches would enable analyses, based on hundreds or thousands of single cells, to reconstruct intracellular circuits, enumerate and redefine cell states and types, and transform our understanding of cellular decision-making on a genomic scale.

METHODS SUMMARY

BMDCs, prepared as previously described¹², were stimulated with LPS for 4 h and then sorted as single cells or populations (10,000 cells) directly into TCL lysis buffer (Qiagen) supplemented with 1% (v/v) 2-mercaptoethanol. After performing a 2.2× clean up with Agencourt RNAClean XP Beads (Beckman Coulter), whole transcriptome-amplified cDNA products were generated using the SMARTer Ultra-low RNA kit (Clontech), and conventional Illumina libraries were made and sequenced to an average depth of 27 million read pairs (HiSeq 2000, Illumina). Expression levels and splicing ratios were quantified using RSEM¹⁴ and MISO¹⁸, respectively. Additional experiments were performed using RNA-FISH (Panomics), immunofluorescence, FACS and single-cell qRT-PCR (Single Cell-to-CT (Invitrogen) and BioMark (Fluidigm)). Full Methods and any associated references are provided in the Supplementary Information.

Received 2 November 2012; accepted 5 April 2013.

Published online 19 May; corrected online 12 June 2013 (see full-text HTML version for details).

- Bengtsson, M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res.* **15**, 1388–1392 (2005).
- Raj, A. & Van Oudenaarden, A. Single-molecule approaches to stochastic gene expression. *Ann. Rev. Biophys.* **38**, 255–270 (2009).
- Kalisky, T., Blainey, P. & Quake, S. R. Genomic analysis at the single-cell level. *Ann. Rev. Gen.* **45**, 431–445 (2011).
- Feinerman, O. *et al.* Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Mol. Sys. Biol.* **6**, 1–16 (2010).
- Cohen, A. A. *et al.* Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**, 1511–1516 (2008).
- Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
- Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).
- Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6**, 377–382 (2009).
- Ramskold, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotech.* **30**, 777–782 (2012).
- Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
- Tay, S. *et al.* Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271 (2010).
- Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257–263 (2009).
- Li, G.-W. & Xie, X. S. Central dogma at the single-molecule level in living cells. *Nature* **475**, 308–315 (2011).
- Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
- Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nature Genet.* **38**, 636–643 (2006).
- Garber, M. *et al.* A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* **47**, 810–822 (2012).
- Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
- Katz, Y., Wang, E. T., Airolidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7**, 1009–1015 (2010).
- Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods* **9**, 72–74 (2012).
- Waks, Z., Klein, A. M. & Silver, P. A. Cell-to-cell variability of alternative RNA splicing. *Mol. Syst. Biol.* **7**, 1–12 (2011).
- Gurskaya, N. G. *et al.* Analysis of alternative splicing of cassette exons at single-cell level using two fluorescent proteins. *Nucleic Acids Res.* **40**, e57 (2012).
- Jiang, A. *et al.* Disruption of E-cadherin-mediated adhesion induces a functionally distinct pathway of dendritic cell maturation. *Immunity* **27**, 610–624 (2007).
- Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805 (2004).
- Ning, S., Huye, L. E. & Pagano, J. S. Regulation of the transcriptional activity of the IRF7 promoter by a pathway independent of interferon signaling. *J. Biol. Chem.* **280**, 12262–12270 (2005).
- Zhao, M., Zhang, J., Phatnani, H., Scheu, S. & Maniatis, T. Stochastic expression of the interferon- β gene. *PLoS Biol.* **10**, e1001249 (2012).
- Rand, U. *et al.* Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Mol. Syst. Biol.* **8**, 584 (2012).
- Änkö, M.-L. *et al.* The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* **13**, R17 (2012).
- Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A. & Nolan, G. P. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**, 523–529 (2005).
- Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
- Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485–490 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Chevrier, C. Villani, M. Jovanovic, M. Bray and J. Shuga for scientific discussions; N. Friedman and E. Lander for comments on the manuscript; B. Tilton, T. Rogers and M. Tam for assistance with cell sorting; J. Bochicchio, E. Shefler and C. Guiducci for project management; the Broad Genomics Platform for all sequencing work; K. Fitzgerald for the *Irf7*^{-/-} bone marrow; and L. Gaffney for help with artwork. Work was supported by a National Institutes of Health (NIH) Postdoctoral Fellowship (1F32HD075541-01, to R.S.), a Charles H. Hood Foundation Postdoctoral Fellowship (to A. Goren), an NIH grant (U54 AI057159, to N.H.), an NIH New Innovator Award (DP2 OD002230, to N.H.), an NIH CEGS Award (1P50HG006193-01, to H.P., A.R. and N.H.), NIH Pioneer Awards (5DP1OD003893-03 to H.P., DP1OD003958-01 to A.R.), the Broad Institute (to H.P. and A.R.), HHMI (to A.R.), and the Klarman Cell Observatory at the Broad Institute (to A.R.).

Author Contributions A.R., H.P., J.Z.L., N.H., A.K.S., R.S., A. Goren and A. Gnirke conceived and designed the study. A.K.S., X.A., R.S.G., J.T.G., R.R., C.M., D.L., J.J.T., D.G. and J.T.G. performed experiments. R.S., A.K.S., S.S. and N.Y. performed computational analyses. R.S., A.K.S., A. Goren, N.H., J.Z.L., H.P. and A.R. wrote the manuscript, with extensive input from all authors.

Author Information Data have been deposited in GEO under accession number GSE41265. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.P. (Hongkun_Park@harvard.edu) or A.R. (aregev@broad.mit.edu).

MBNL proteins repress ES-cell-specific alternative splicing and reprogramming

Hong Han^{1,2*}, Manuel Irimia^{1*}, P. Joel Ross³, Hoon-Ki Sung⁴, Babak Alipanahi⁵, Laurent David⁶, Azadeh Golipour^{2,6}, Mathieu Gabut¹, Iacovos P. Michael⁴, Emil N. Nachman^{1,2}, Eric Wang⁷, Dan Trcka⁶, Tadeo Thompson³, Dave O'Hanlon¹, Valentina Slobodeniuc¹, Nuno L. Barbosa-Morais^{1,8}, Christopher B. Burge⁷, Jason Moffat^{1,2}, Brendan J. Frey^{1,5}, Andras Nagy^{4,9}, James Ellis^{2,3}, Jeffrey L. Wrana^{2,6} & Benjamin J. Blencowe^{1,2}

Previous investigations of the core gene regulatory circuitry that controls the pluripotency of embryonic stem (ES) cells have largely focused on the roles of transcription, chromatin and non-coding RNA regulators^{1–3}. Alternative splicing represents a widely acting mode of gene regulation^{4–8}, yet its role in regulating ES-cell pluripotency and differentiation is poorly understood. Here we identify the muscleblind-like RNA binding proteins, MBNL1 and MBNL2, as conserved and direct negative regulators of a large program of cassette exon alternative splicing events that are differentially regulated between ES cells and other cell types. Knockdown of MBNL proteins in differentiated cells causes switching to an ES-cell-like alternative splicing pattern for approximately half of these events, whereas overexpression of MBNL proteins in ES cells promotes differentiated-cell-like alternative splicing patterns. Among the MBNL-regulated events is an ES-cell-specific alternative splicing switch in the forkhead family transcription factor FOXP1 that controls pluripotency⁹. Consistent with a central and negative regulatory role for MBNL proteins in pluripotency, their knockdown significantly enhances the expression of key pluripotency genes and the formation of induced pluripotent stem cells during somatic cell reprogramming.

A core set of transcription factors that includes OCT4 (also called POU5F1), NANOG and SOX2, together with specific microRNAs and long non-coding RNAs, control the expression of genes required for the establishment and maintenance of ES-cell pluripotency^{1–3,10–12}. Alternative splicing, the process by which splice sites in primary transcripts are differentially selected to produce structurally and functionally distinct messenger RNA and protein isoforms, provides a powerful additional mechanism with which to control cell fate^{7,8,13}, yet its role in the regulation of pluripotency has only recently begun to emerge. In particular, the inclusion of a highly conserved ES-cell-specific 'switch' exon in the FOXP1 transcription factor changes its DNA binding specificity such that it stimulates the expression of pluripotency transcription factors, including OCT4 and NANOG, while repressing genes required for differentiation⁹. However, the *trans*-acting regulators of this and other alternative splicing events^{14–16} implicated in ES-cell biology are not known. These factors are important to identify, as they may control regulatory cascades that direct cell fate, and likewise they may also control the efficiency and kinetics of somatic cell reprogramming.

To identify such factors, we used high-throughput RNA sequencing (RNA-seq) data to define human and mouse cassette alternative exons that are differentially spliced between ES cells and induced pluripotent stem cells (iPSCs), and diverse differentiated cells and tissues, referred to below as 'ES-cell-differential alternative splicing'. A splicing code

analysis¹⁷ was then performed to identify *cis*-elements that may promote or repress these exons. The RNA-seq data used to profile alternative splicing were also used to detect human and mouse splicing factor genes that are differentially expressed between ES cells/iPSCs and non-ES cells/tissues. By integrating these data sources, we sought to identify differentially expressed splicing regulators with defined binding sites that match *cis*-elements predicted by the code analysis to function in ES-cell-differential alternative splicing.

We identified 181 human and 103 mouse ES-cell-differential alternative splicing events, with comparable proportions of exons that are $\geq 25\%$ more included or more skipped in ES cells versus the other profiled cells and tissues (Fig. 1a, Supplementary Figs 1a and 2, and Supplementary Tables 1 and 2). When comparing orthologous exons in both species, 25 of the human and mouse ES-cell-differential alternative splicing events overlapped ($P < 2.2 \times 10^{-16}$; hypergeometric test). The human and mouse ES-cell-differential alternative splicing events are significantly enriched in genes associated with the cytoskeleton (for example, *DST*, *ADD3*), plasma membrane (for example, *DNM2*, *ITGA6*) and kinase activity (for example, *CASK*, *MARK2* and *MAP2K7*) (Supplementary Table 3). They also include the aforementioned FOXP1 ES-cell-switch alternative splicing event, and previously unknown alternative splicing events in other transcription or chromatin regulatory factor genes (for example, *TEAD1* and *MTA1*) that have been implicated in controlling pluripotency^{18,19}. These results suggest a considerably more extensive role for regulated alternative splicing in ES-cell biology than previously appreciated.

The splicing code analysis revealed that motifs corresponding to consensus binding sites of the conserved MBNL proteins are the most strongly associated with ES-cell-differential alternative splicing in human and mouse. The presence of MBNL motifs in downstream flanking intronic sequences is associated with exon skipping in ES cells, whereas their presence in upstream flanking intronic sequences is associated with exon inclusion in ES cells (Fig. 1b, human code; Supplementary Fig. 1b, mouse code). To a lesser extent, features resembling binding sites for other splicing regulators, including polypyrimidine tract binding protein (PTBP) and RNA-binding fox (RBFOX) proteins, may also be associated with ES-cell-differential alternative splicing.

From RNA-seq expression profiling of 221 known or putative splicing factors, 11 genes showed significant differential expression between human ES cells/iPSCs and other cells and tissues (Bonferroni-corrected $P < 0.05$, Wilcoxon rank-sum test) (Fig. 1c and Supplementary Table 4). Notably, *MBNL1* and *MBNL2* had the lowest relative mRNA levels in ES cells/iPSCs compared to other cells and tissues

¹Banting and Best Department of Medical Research and Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada. ²Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ³Developmental and Stem Cell Biology, The Hospital for Sick Children, 101 College Street, Toronto, Ontario M5G 1L7, Canada. ⁴Center for Stem Cells and Tissue Engineering, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. ⁵Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 3G4, Canada. ⁶Center for Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. ⁷Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ⁸Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal. ⁹Department of Obstetrics and Gynecology, University of Toronto, Toronto, Ontario, M5S 1A8, Canada.

*These authors contributed equally to this work.

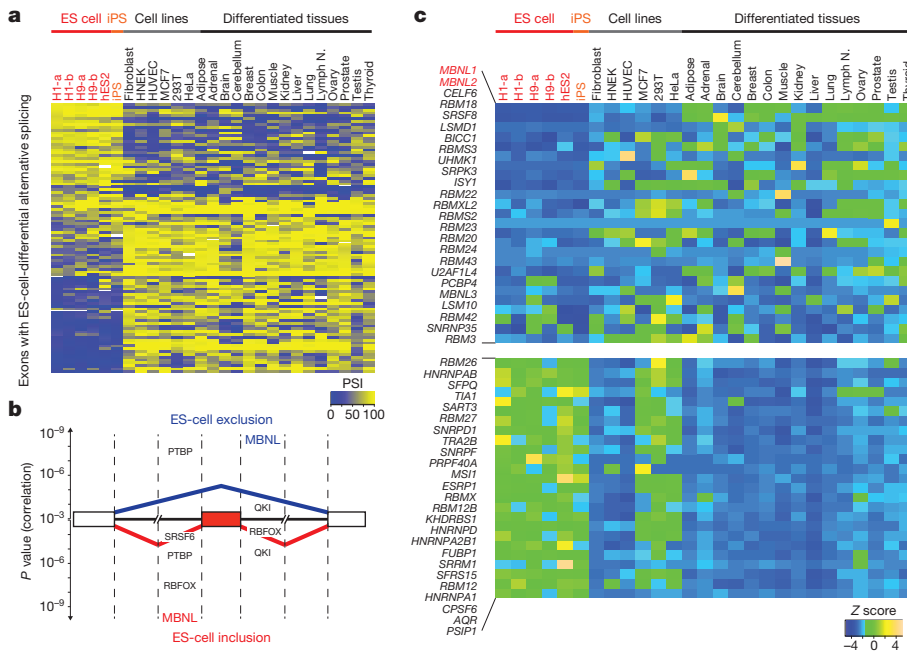


Figure 1 | Identification of regulators of ES-cell-differential alternative splicing. **a**, Heat map of per cent spliced in (PSI) values for 95 representative ES-cell-differential alternative splicing events in transcripts that are widely expressed across human ES cells/iPSCs, non-ES-cell lines and differentiated tissues. **b**, Splicing code features that are significantly associated with ES-cell-differential alternative splicing. Features are ranked according to Pearson correlation *P* values (*y* axis) for alternative exons with either lower (top) or

higher (bottom) inclusion in ES cells/iPSCs. Dashed lines indicate 300 nucleotide intervals from splice sites. **c**, Heat map of Z scores of mRNA expression (cRPKM) levels for splicing factors. Twenty-five splicing factors with the lowest or highest relative mRNA expression levels in ES cells/iPSCs compared to non-ES cells/tissues are shown. cRPKM, corrected reads per kilobase exon model per million reads³⁰.

(Fig. 1c, Supplementary Fig. 3a and Methods). Quantitative RT-PCR (polymerase chain reaction with reverse transcription; qRT-PCR) assays confirmed this observation (Supplementary Fig. 3b). Similar results were obtained when analysing mouse expression data (Supplementary Fig. 3c–e and Supplementary Table 4). PTBP, RBFOX and other splicing factors potentially associated with ES-cell-differential alternative splicing by the splicing code analysis did not exhibit significant differences in mRNA levels between ES cells/iPSCs and other cells or tissues. Collectively, these results suggest a conserved and prominent role for MBNL1 and MBNL2 in ES-cell-differential alternative splicing.

Because MBNL proteins are expressed at minimal levels in ES cells compared to other cell types, we proposed that they may repress ES-cell-differential exons in non-ES cells, and/or activate the inclusion of exons in non-ES cells that are skipped in ES cells. Indeed, previous studies have shown that in differentiated cells, MBNL proteins suppress exon inclusion when they bind upstream flanking intronic sequences, and they promote inclusion when binding to downstream flanking intronic sequences^{20,21}. The results of the splicing code analysis are consistent with this mode of regulation, when taking into account that MBNL proteins are depleted in ES cells relative to differentiated cells and tissues (Fig. 1b and Supplementary Fig. 1b).

To test the above hypothesis, we used short interfering RNAs (siRNAs) to knock down MBNL1 and MBNL2 (to ~10% of their endogenous levels), individually or together, in human (293T and HeLa) and mouse (neuro2A (N2A)) cells (Fig. 2a and Supplementary Fig. 4a; see below). For comparison, knockdowns were performed in human (H9) and mouse (CGR8) ES cells. RT-PCR assays were used to monitor the ES-cell-switch exon of *FOXP1/Foxp1* (human exon 18b/mouse exon 16b), which is partially included in ES cells and fully skipped in differentiated cell types⁹. The splicing code analysis suggested that this exon is associated with conserved regulation by MBNL proteins, through possible direct disruption of splice-site recognition (Fig. 2b; see legend and below). Knockdown of MBNL2 in 293T or HeLa cells resulted in a <1% increase in *FOXP1* exon 18b inclusion, whereas

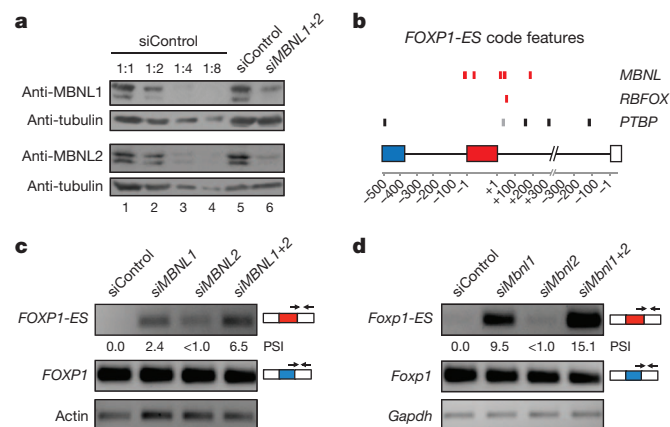


Figure 2 | MBNL proteins regulate ES-cell-specific alternative splicing. **a**, Western blots confirming efficient knockdown of MBNL1 and MBNL2 proteins in human 293T cells transfected with siRNA pools targeting these factors (*siMBNL1*+2, lane 6). Lane 5, lysate from cells transfected with a non-targeting siRNA pool (siControl). Lanes 1–4, serial dilutions (1:1, 1:2, 1:4 and 1:8) of lysate from cells transfected with siControl. **b**, Splicing code map highlighting genomic locations of *MBNL*, *RBFox* and *PTBP* motifs associated with ES-cell-specific alternative splicing of *FOXP1*/*Foxp1* exon 18b/16b, the inclusion of which forms the *FOXP1-ES*/*Foxp1-ES* isoform. Human (black), mouse (grey) or conserved features (red) are indicated. Note that conserved MBNL motifs are associated with possible direct interference of exon 18b/16b splice site regulation. **c**, RT–PCR assays monitoring mRNA levels of *FOXP1* canonical (blue exon) and *FOXP1-ES* (red exon) isoforms in 293T cells transfected with siControl, *siMBNL1*, *siMBNL2* or *siMBNL1*+2. RT–PCR used splice-junction-specific primers, as indicated. Expression levels of actin are shown as loading controls. **d**, mRNA levels of murine *Foxp1*-canonical and *Foxp1-ES* isoforms were assayed as in **c** in N2A cells. Expression levels of *Gapdh* are shown as loading controls.

knockdown of MBNL1 alone, or together with MBNL2, resulted in increases in PSI (per cent spliced in), from zero to 2–2.4% and 6.5–7.1%, respectively (Figs 2c and 3c and Supplementary Figs 4b and 5). More pronounced effects were observed for *Foxp1* exon 16b in N2A cells (PSI shift from 0 to 15.1 for the double knockdown; Fig. 2d and Supplementary Fig. 4c). Knockdowns in ES cells had modest effects on exon 18b/16b splicing, consistent with the low levels of MBNL expression in these cells (Supplementary Fig. 4d, e). Knockdown of a third MBNL family member, MBNL3, which has a more restricted cell-type distribution compared to MBNL1 and MBNL2 (ref. 22), had no detectable effect on exon 18b splicing (Supplementary Fig. 5). These results suggest that MBNL1 and MBNL2 proteins have conserved and partially redundant roles in the negative regulation of *FOXP1*/*Foxp1* exon 18b/16b inclusion.

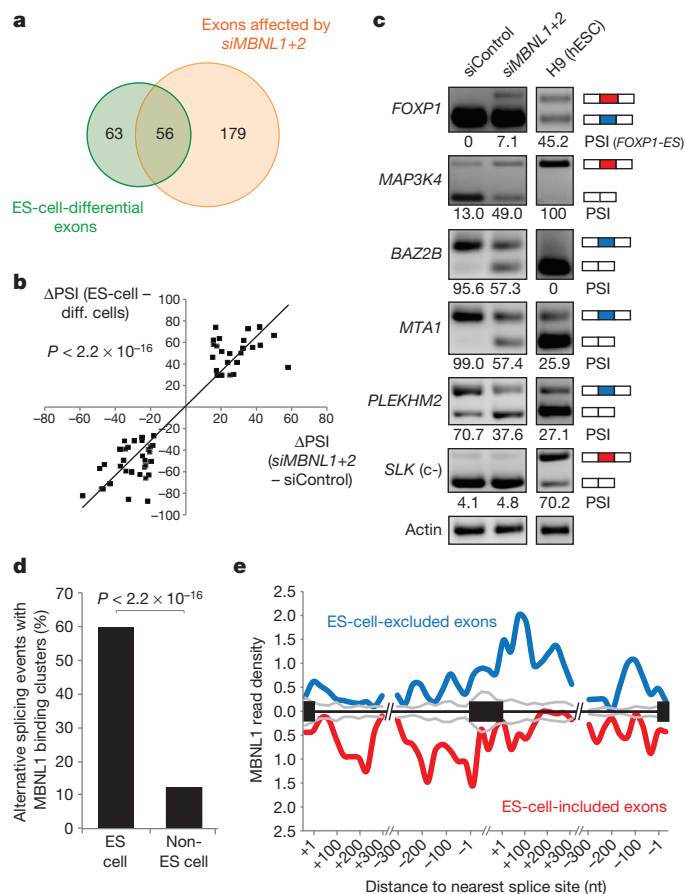


Figure 3 | MBNL proteins regulate approximately half of ES-cell-differential alternative splicing events. **a**, Venn diagram showing the proportion of ES-cell-differential alternative splicing events (green) that display ≥ 15 PSI change between HeLa cells transfected with siRNA pools targeting MBNL1 and MBNL2 (siMBNL1+2) versus siControl pool (orange). **b**, High association ($P < 2.2 \times 10^{-16}$, one-sided binomial test between quadrants) between differences in PSIs of ES cells versus differentiated cells/tissues, and differences in PSIs of siMBNL1+2 knockdown versus siControl treatments. **c**, Representative RT-PCR validations for ES-cell-differential alternative splicing events that have PSI changes in HeLa cells after siMBNL1+2 transfection and for ES-cell-differential alternative splicing events that do not change upon siMBNL1+2 knockdown (c-); splicing patterns in human H9 ES cells are shown for comparison. **d**, Percentage of alternative splicing events with overlapping MBNL1 CLIP-seq binding clusters²⁰ in C2C12 cells for ES-cell-differential or non-ES-cell-regulated alternative exons. **e**, Merged map of MBNL binding clusters in transcripts with ES-cell-differential alternative splicing events. Maps of MBNL1 binding sites with respect to exons that have higher or lower inclusion in ES cells/iPSCs, relative to non-ES cells/tissues.

To assess the extent to which ES-cell-differential alternative splicing events are controlled by MBNL proteins, MBNL1 and MBNL2 were knocked down in HeLa cells, and RNA-seq profiling was used to detect alternative splicing changes (Fig. 3). Of 119 profiled ES-cell-differentially spliced exons, nearly half are affected by knockdown of MBNL proteins, with a ≥ 15 PSI change towards an ES-cell-like alternative splicing pattern (Fig. 3a). A strong overall association ($P < 2.2 \times 10^{-16}$, one-sided binomial test) was observed when comparing PSI changes for exons differentially spliced between ES cells and non-ES cells/tissues, and PSI changes for the same exons following knockdown of MBNL proteins (Fig. 3b). RT-PCR experiments confirmed all analysed MBNL knockdown-dependent and -independent PSI changes (Fig. 3c and Supplementary Fig. 6a). The specificity of the knockdown experiments was further demonstrated by comparing individual siRNAs that target different sequences within MBNL1 transcripts (Supplementary Fig. 6b). Comparable results were observed when MBNL1 and MBNL2 proteins were simultaneously knocked down in 293T cells, and in undifferentiated C2C12 mouse myoblast cells (Supplementary Fig. 7). Conversely, overexpression of MBNL1 and MBNL2 proteins in mouse ES cells promoted differentiated-cell-like patterns for all analysed ES-cell-differential alternative splicing events (Supplementary Fig. 8), including a switch to the exclusive use of the canonical (that is, non-ES cell) exon 16 in *Foxp1* transcripts. Consistent with this observation, overexpression of MBNL proteins in ES cells also led to increased kinetics of silencing of core pluripotency factors upon differentiation, and further promoted the expression of specific lineage markers representative of all three germ layers (Supplementary Fig. 9).

Mapping of MBNL protein binding to endogenous transcripts using ultraviolet crosslinking coupled to immunoprecipitation and sequencing (CLIP-seq or HITS-CLIP²³) in undifferentiated C2C12 myoblast cells²⁰ confirmed that these proteins directly target ES-cell-differential alternative splicing events, including *Foxp1* exon 16b (Fig. 3d and Supplementary Fig. 10a). Of 57 mouse ES-cell-differential exons expressed in C2C12 cells, ~34 (60%) are associated with overlapping or proximal clusters of MBNL CLIP-seq tags ('binding clusters'), whereas binding clusters are associated with 72 out of 601 (12%) of exons that are not differentially regulated in ES cells ($P < 2.2 \times 10^{-16}$, proportion test; Fig. 3d). The binding clusters associated with ES-cell-differential alternative splicing are significantly enriched in consensus binding sites for MBNL proteins (Supplementary Fig. 10b)^{20,21,24}. Moreover, consistent with the splicing code analysis (Fig. 1b and Supplementary Fig. 1b) and previous results^{20,21}, the locations of MBNL binding clusters correlate with whether the target exons are more or less included in ES cells compared to other cells and tissues (Fig. 3e). Collectively, the results so far demonstrate that MBNL proteins act widely and directly to regulate ES-cell-differential alternative splicing, and consequently pluripotency factor expression.

We next asked whether MBNL proteins have an impact on somatic cell reprogramming (Fig. 4a). Secondary mouse embryonic fibroblasts (MEFs)²⁵ expressing the 'OKSM' transcription factors (OCT4, KLF4, SOX2, c-MYC)²⁶ from transgenes under doxycycline-inducible control were transfected with siRNA pools to knock down MBNL1 and MBNL2 (siMBNL1+2), or with a control, non-targeting siRNA pool (siControl). At days 3 and 5 after doxycycline induction, mRNA expression of endogenous pluripotency genes, including *Oct4*, *Nanog*, *Sall4* and *Alpl*, were assayed by qRT-PCR (Fig. 4b and Supplementary Fig. 11a). None of these genes displayed significant changes in expression at day 3 (Fig. 4a, b); however, at day 5, MBNL knockdown stimulated their expression by approximately twofold over the siControl treatment (Fig. 4b and Supplementary Fig. 11a). MBNL knockdown also resulted in a ~30% increase in the colony area immunostained for SSEA1, a pluripotency-associated marker expressed early during reprogramming (Fig. 4c). In contrast, knockdown of OCT4 (siOCT4) resulted in significant reductions in endogenous pluripotency gene expression and in SSEA1-positive colonies (Fig. 4b, c).

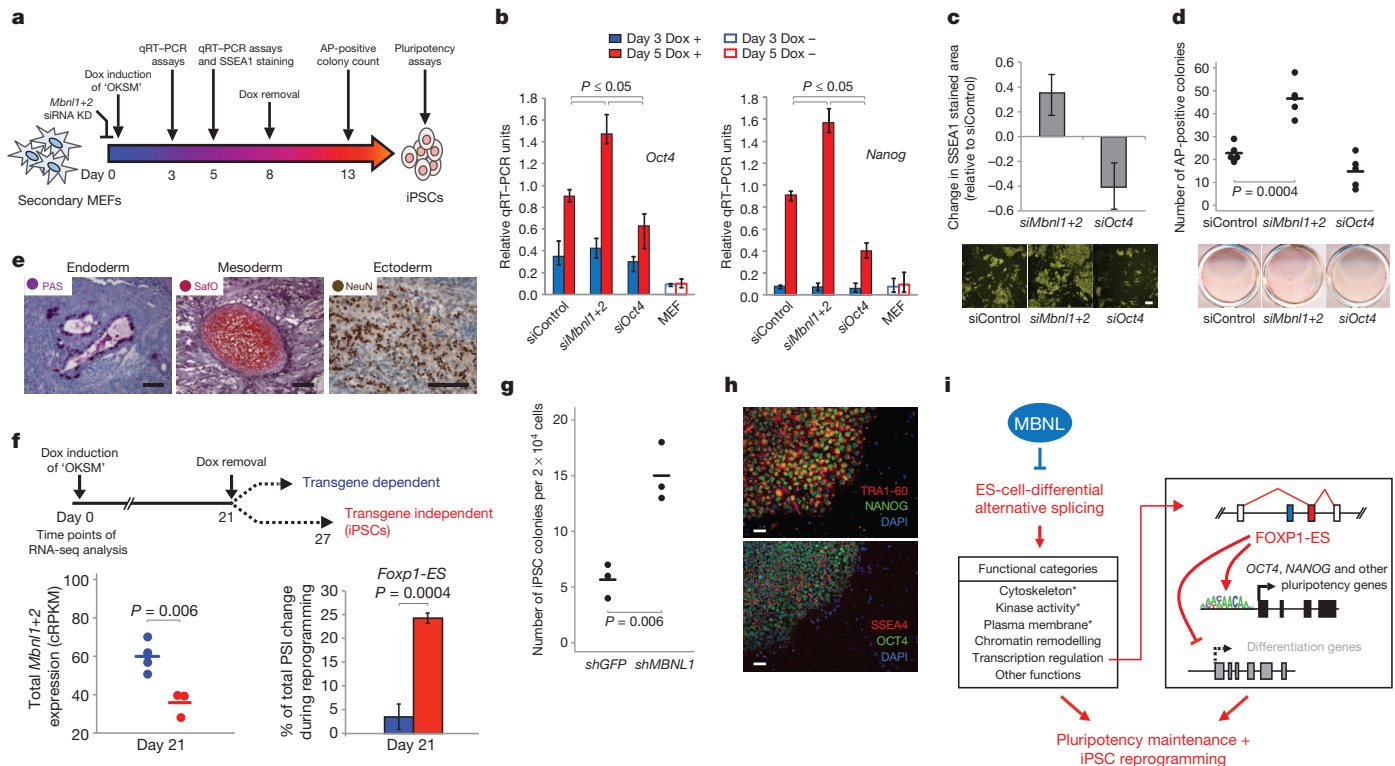


Figure 4 | Knockdown of MBNL proteins enhances reprogramming efficiency and kinetics. **a**, Experimental scheme. **b**, qRT-PCR quantification of mRNA expression levels of endogenous *Oct4* and *Nanog* (data for additional genes in Supplementary Fig. 11a). Secondary MEFs were transfected with control siRNAs (siControl), siRNAs targeting *Mbnl1* and *Mbnl2* (si*Mbnl1+2*) or *Oct4* (si*Oct4*) and treated with doxycycline (Dox) for 3 days (blue bars) or 5 days (red bars) before analysis. Empty bars, secondary MEFs without doxycycline induction. Values represent means \pm range ($n = 3$). **c**, Top: quantification of SSEA1-stained area change relative to siControl at day 5 after doxycycline induction; values represent means \pm range ($n = 3$). Bottom: representative images of SSEA1 staining. Scale bar, 100 μ m. **d**, Top: quantification of doxycycline-independent iPSC colony formation. Secondary MEFs were treated with doxycycline for 8 days followed by 5 days of doxycycline withdrawal and counting of alkaline-phosphatase-positive colonies. Bottom: representative images of alkaline phosphatase staining. **e**, Teratoma assay assessing the pluripotency potential of iPSCs derived from secondary MEFs after knockdown of MBNL proteins. Haematoxylin and eosin staining, with additional staining/immunolabelling using periodic acid-Schiff (PAS; for detection of glycogen or glycoprotein producing cells), safranin O (Safo; for detection of cartilage), or

Successful reprogramming requires that cells undergo a transition to an OKSM transgene-independent state²⁷. We therefore asked whether suppression of MBNL proteins promotes transgene independence. OKSM transgenes were induced with doxycycline for 8 days, then the cells were cultured for 5 days without doxycycline (Fig. 4a). Whereas knockdown of OCT4 reduced colony formation, knockdown of MBNL proteins resulted in an approximate twofold increase in transgene-independent colonies, as detected by alkaline phosphatase staining ($P = 0.0004$; one-sided t -test) (Fig. 4d and Supplementary Fig. 11b). iPSC lines derived from transgene-independent colonies after MBNL knockdown were pluripotent and contributed to all three germ layers in both teratoma and chimera assays (Fig. 4e and Supplementary Figs 11–13). Consistent with these results, *Mbnl* expression is significantly reduced in secondary MEF clones²⁷ cultured in the presence of doxycycline that are competent to achieve transgene independence (when doxycycline is removed) versus those that are not ($P = 0.006$; one-sided t -test) (Fig. 4f, left). Moreover, the PSI levels of ES-cell-differential alternative splicing events, including *Foxp1* exon 16b, significantly

correlate with ES cell/iPSC alternative splicing patterns only in those clones that are competent to transition to transgene independence (Fig. 4f, right; Supplementary Fig. 14 and Supplementary Table 5; $r = 0.80$, $P = 3.2 \times 10^{-12}$). Notably, knockdown of MBNL1 in human fibroblasts expressing OKSM also resulted in an approximate twofold increase in the appearance of iPSC colonies (Fig. 4g, h and Supplementary Fig. 15). MBNL proteins thus have a conserved, negative regulatory role in somatic cell reprogramming.

The results of this study reveal that MBNL proteins negatively regulate an ES-cell-differential alternative splicing network that controls pluripotency and reprogramming (Fig. 4i). These proteins probably act, in part, by directly repressing the ES-cell-specific splicing switch in *FOXP1*, which promotes the expression of core pluripotency genes. However, additional genes with MBNL-regulated alternative splicing events have been linked to the control of pluripotency, indicating a more extensive role for the alternative splicing network in ES-cell biology (Fig. 4i). These observations represent the first evidence that *trans*-acting splicing regulators have a central role in the core circuitry

required for ES-cell pluripotency and reprogramming. Our results further offer a potential new approach for enhancing the production of iPSCs for research and therapeutic applications.

METHODS SUMMARY

siRNA knockdown and RNA analysis. Cells were transfected with SMART-pool siRNAs (Dharmacon) using DharmaFECT1 reagent and collected 48 or 72 h after transfection. Secondary MEFs were transfected with siRNA pools using Lipofectamine RNAiMAX (Invitrogen), as described previously²⁸, and OKSM transgenes were induced 24 h after with doxycycline. Semi-quantitative RT-PCR assays were performed using the OneStep RT-PCR kit (Qiagen), with modifications as described previously²⁹. Quantitative RT-PCR assays were performed as previously described⁹. Primer sequences are available upon request.

iPSC colony formation assays and characterization. Cells were transfected with siRNAs pools and treated with doxycycline for 5 days before imaging using an IN Cell Analyzer 2000 (GE Healthcare). To assay formation of doxycycline-independent colonies, secondary MEFs transfected with siRNA pools were treated with doxycycline for 8 days. Cell counting was performed before and at each passage after siRNA transfection. At day 8, the same number of cells were passaged in doxycycline-free media in 12-well plates and cultured until day 13, when they were fixed and stained with alkaline phosphatase for colony counting. For single-cell assays, secondary MEFs were induced by doxycycline treatment and clonal derivatives were cultured for 21 days. Removal of doxycycline at day 21 revealed alkaline-phosphatase-positive colonies (transgene-independent clones) and failed colonies (transgene-dependent clones). RNA-seq analysis was performed on three transgene-independent and five transgene-dependent clones at day 21 after doxycycline induction, as previously described²⁷. Details of human iPSC generation and characterization are available in the Methods.

Teratoma and chimaera analysis. ES cells were injected subcutaneously into dorsal flanks of nude mice (CByJ.Cg-Foxn1nu/J) and resulting teratomas were analysed using immunohistochemistry or cell-specific staining 4 to 5 weeks after injection. Chimaera aggregation and whole-mount staining were performed as previously described²⁷.

Full Methods and any associated references are available in the online version of the paper.

Received 12 September 2012; accepted 7 May 2013.

Published online 5 June 2013.

- Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
- Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
- Bao, X. *et al.* MicroRNAs in somatic cell reprogramming. *Curr. Opin. Cell Biol.* **25**, 208–214 (2013).
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).
- Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
- Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252–1269 (2013).
- Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–463 (2010).
- Kalsotra, A. & Cooper, T. A. Functional consequences of developmentally regulated alternative splicing. *Nature Rev. Genet.* **12**, 715–729 (2011).
- Gabut, M. *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147**, 132–146 (2011).
- Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
- Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).
- Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722–737 (2009).
- Irimia, M. & Blencowe, B. J. Alternative splicing: decoding an expansive regulatory layer. *Curr. Opin. Cell Biol.* **24**, 323–332 (2012).

- Rao, S. *et al.* Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. *Mol. Cell. Biol.* **30**, 5364–5380 (2010).
- Salomonis, N. *et al.* Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc. Natl Acad. Sci. USA* **107**, 10514–10519 (2010).
- Mayshar, Y. *et al.* Fibroblast growth factor 4 and its novel splice isoform have opposing effects on the maintenance of human embryonic stem cell self-renewal. *Stem Cells* **26**, 767–774 (2008).
- Barash, Y. *et al.* Deciphering the splicing code. *Nature* **465**, 53–59 (2010).
- Liang, J. *et al.* Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nature Cell Biol.* **10**, 731–739 (2008).
- Lian, I. *et al.* The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev.* **24**, 1106–1118 (2010).
- Wang, E. T. *et al.* Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150**, 710–724 (2012).
- Charizanis, K. *et al.* Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75**, 437–450 (2012).
- Pascual, M., Vicente, M., Monferrer, L. & Artero, R. The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* **74**, 65–80 (2006).
- Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
- Fernandez-Costa, J. M., Llamusi, M. B., Garcia-Lopez, A. & Artero, R. Alternative splicing regulation by Muscleblind proteins: from development to disease. *Biol. Rev. Camb. Philos. Soc.* **86**, 947–958 (2011).
- Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766–770 (2009).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11**, 769–782 (2012).
- Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
- Calarco, J. A. *et al.* Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* **21**, 2963–2975 (2007).
- Labbé, R. M. *et al.* A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**, 1734–1745 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank U. Braunschweig, J. Ellis, S. Gueroussov and B. Raj for comments on the manuscript. We acknowledge D. Torti in the Donnelly Sequencing Centre for sequencing samples; L. Lee for assisting with the splicing code analysis; J. Garner (Hospital for Sick Children Embryonic Stem Cell Facility) for preparing feeder cells; A. Piekna for morphological examination of human iPSC colonies; M. Narimatsu for assisting with chimaerism analysis; and P. Mero for assisting with cell imaging. This work was supported by grants from the Canadian Institutes of Health Research (CIHR) (to B.J.B., J.L.W., A.N., J.E. and B.J.F.), the Ontario Research Fund (to J.L.W., B.J.B., A.N. and others), the Canadian Stem Cell Network (to A.N. and B.J.B.), and by a grant from the National Institutes of Health (R33MH087908) to J.E. H.H. was supported by a University of Toronto Open Fellowship. P.J.R., M.I. and N.L.B.-M. were supported by postdoctoral fellowships from the Ontario Stem Cell Initiative, Human Frontiers Science Program Organization, and the Marie Curie Actions, respectively.

Author Contributions H.H. performed experiments in Figs 1–4 and Supplementary Figs 2–9 and 11–13. M.I. performed bioinformatic analyses in Figs 1–4 and Supplementary Figs 1, 3, 7, 10 and 14, with input from N.L.B.-M. L.D. and A.G. assisted with secondary MEF reprogramming experiments and clone characterization, and D.T. generated secondary MEF lines and performed chimaerism testing. P.J.R., T.T. and M.G. performed human reprogramming experiments and iPSC characterization. H.K.S. performed teratoma assays. B.A. and B.J.F. generated splicing code data. I.P.M., H.-K.S. and D.O. assisted with ES-cell overexpression and differentiation experiments. E.W. and C.B.B. generated and analysed CLIP-seq data. E.N.N. and V.S. performed RT-PCR validation experiments. B.J.B., H.H. and M.I. designed the study, with input from J.L.W., J.E., A.N. and J.M. B.J.B., H.H. and M.I. wrote the manuscript, with input from the other authors.

Author Information GEO accession numbers are provided in Supplementary Table 1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.J.B. (b.blencowe@utoronto.ca).

METHODS

siRNA knockdown and RNA analysis. Cells were transfected with SMART-pool siRNAs (Dharmacon) using DharmaFECT1 reagent and collected 48 or 72 h after transfection. Secondary MEFs were transfected with siRNA pools using Lipofectamine RNAiMAX (Invitrogen), as described previously²⁸, and OKSM transgenes were induced 24 h after with doxycycline. Semi-quantitative RT-PCR assays were performed using the OneStep RT-PCR kit (Qiagen), with modifications as described previously²⁹. Quantitative RT-PCR (qRT-PCR) assays were performed as previously described⁹. Primer sequences are available upon request.

Cell lines and cell culture. HeLa, 293T and C2C12 cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and antibiotics (penicillin/streptomycin). Neuro2A (N2A) cells were grown in DMEM supplemented with 10% FBS, sodium pyruvate, MEM non-essential amino acids, and penicillin/streptomycin. H9 human ES cells, CGR8 and R1 mouse ES cells were cultured as described previously³¹. Secondary mouse embryonic fibroblasts (MEFs) were maintained in DMEM supplemented with 10% FBS, L-glutamine and penicillin/streptomycin on 0.1% gelatin-coated plates. During reprogramming, secondary MEFs were grown in mouse ES media and induced to express OKSM factors using $1.5 \mu\text{g ml}^{-1}$ of doxycycline as described previously^{32,33}.

Protein extraction and western blotting. Cell pellets were lysed in radio-immunoprecipitation assay (RIPA) buffer by brief sonication. Protein lysate (30–150 μg) was separated on a 10% SDS–polyacrylamide gel and transferred to a PVDF membrane. The membranes were blotted with the following antibodies: anti-Flag M2 (1:1500, Sigma), anti-MBNL1 (1:500, Abcam), anti-MBNL2 (1:200, Santa Cruz Biotechnology) and anti- α -tubulin (1:5000, Sigma). Secondary antibodies (GE Healthcare) and chemiluminescence reagents (Perkin Elmer) were used as per the manufacturer's instructions.

RNA extraction and qRT-PCR assays. Total RNA was extracted using TRI Reagent (Sigma) or RNeasy columns (Qiagen). RT-PCR assays were performed using the OneStep RT-PCR kit (Qiagen), as per the manufacturer's instructions. 20 ng total RNA or 1 ng of polyA+ RNA was used per 10- μl reaction. Radiolabelled reactions contained 0.3 μCi of α -³²P-dCTP per 10- μl reaction. The number of amplification cycles was 22 for actin and *Gapdh*, and 27–32 for all other transcripts analysed. Reaction products were separated on 1–3% agarose gels. Quantification of isoform abundance was performed using either ImageQuant (GE Healthcare) or ImageJ software. To amplify the *FOXP1/Foxp1* isoforms selectively (Fig. 2c, d), primers specific for splice junctions were used.

For quantitative RT-PCR, first-strand cDNAs were generated from 1–3 μg of total RNA or 100 ng of polyA+ RNA using SuperScript III Reverse Transcriptase (Invitrogen), as per the manufacturer's recommendations, and diluted to 20 ng μl^{-1} and 1 ng μl^{-1} , respectively. qPCR reactions were performed in a 384-well format using 1 μl of each diluted cDNA and FastStart Universal SYBR Green Master (Roche Applied Science). All primers sequences are available upon request.

Immunofluorescence. For immunofluorescence experiments, cells were fixed in 4% PFA for 10 min at room temperature, washed with PBS, and permeabilized for 10 min at 4 °C with 0.1% Triton X-100. After 1 h of blocking, cells were incubated with primary antibodies overnight at 4 °C, and then with secondary antibodies for 1 h at room temperature. Nuclei were stained with Hoechst 33258 (1:5,000, Sigma-Aldrich). Primary antibodies used in this study are: mouse IgM anti-SSEA1 (1:500, BD Biosciences), mouse anti-OCT4 (1:200, BD Biosciences), rabbit anti-NANOG (1:200, Cosmo Bio), goat anti-DPPA4 (1:250, R&D), mouse IgM anti-TRA1-60 (1:100, Invitrogen), rabbit anti-NANOG (1:400, Cell Signaling), mouse anti-SSEA4 (1:100, Invitrogen), rabbit anti-OCT4 (1:200, Abcam), mouse anti-alpha fetoprotein (1:200, R&D), mouse anti-smooth muscle actin (1:200, Invitrogen), and mouse anti-beta-III-tubulin (1:200, Millipore). Secondary antibodies used in this study are: anti-mouse IgM Alexa555 (1:1000, Molecular Probes), anti-mouse IgG Alexa555 (1:1,000, Molecular Probes), anti-rabbit IgG Alexa594 (1:1,000, Molecular Probes), anti-rabbit IgG Alexa488 (1:500, Molecular Probes), and anti-goat IgG Alexa546 (1:1,000, Molecular Probes).

iPSC colony-formation assays and imaging from secondary MEF reprogramming. Secondary MEFs were seeded in 12-well plates, transfected with siRNA pools, and treated with doxycycline for 5 days before fixing and staining. The plates were imaged (for both SSEA1-immunostained and DAPI channels) using an IN Cell Analyzer 2000 (GE Healthcare) with a $\times 4$ objective. For each well, 20 non-overlapping fields were captured and images were analysed using the Columbus System (PerkinElmer). A custom script was generated to identify SSEA1-positive and DAPI-positive colonies. The overall signal in each well was determined using the sum of the overlap area for the 20 fields captured.

To assay the formation of doxycycline-independent colonies, secondary MEFs transfected with siRNA pools were treated with doxycycline for 8 days. Cell counting was performed before and at each passage after siRNA transfection and doubling rates were determined not to change significantly (data not shown). At day 8,

the same number of cells were passaged into doxycycline-free mES-cell media on 12-well plates and cultured until day 13, when they were fixed and stained with alkaline phosphatase for colony counting.

Teratoma analysis. Cells were suspended in PBS and Matrigel (BD Bioscience) mixed solution, and 1×10^6 cells in 100 μl were injected subcutaneously into both dorsal flanks of nude mice (CByJ.Cg-Foxn1nu/J) anaesthetized with isoflurane. Four to five weeks after injection, mice were killed and teratomas were dissected, fixed overnight in 10% buffered formalin phosphate, and embedded in paraffin. Three-to-four-micrometre-thick sections were deparaffinized and hydrated in distilled water. Sections were stained either with haematoxylin and eosin for regular histological examination, or with the following dyes: 0.1% safranin O solution (cartilage, mesoderm-derived tissue) or 0.5% PAS solution (glycoprotein-producing intestinal cell, endoderm-derived tissue). For immunohistochemistry, sections were deparaffinized and hydrated, and antigen retrieval process was performed. After blocking, sections were incubated overnight at 4 °C with primary monoclonal antibody (1:100, Millipore MAB377, clone A60) specific for neuronal nuclear antigen (NeuN, ectoderm-derived tissue), followed by washing in PBS. After 1 h of incubation with secondary anti-mouse-HRP conjugated antibody (1:500, Jackson ImmunoResearch, 115-035-003), signal was visualized by DAB (3,3'-diaminobenzidine; Vector Laboratories, SK-4100) substrate for 5–20 min. Sections were counter-stained with haematoxylin.

Chimaerism analysis. Chimaera aggregation and whole-mount staining were performed as described previously³⁴. Chimeras were obtained through aggregation of siMbnl iPSC clumps with diploid Hsd:ICR(CD-1) embryos. E10.5 embryos were dissected after doxycycline treatment *in utero* via ingestion 24 h before dissection. After dissection, embryos were fixed with 0.25% glutaraldehyde, rinsed in wash buffer (2 mM MgCl₂, 0.01% sodium deoxycholate, and 0.02% Nonidet-P40 in PBS) and then stained overnight in LacZ staining solution (20 mM MgCl₂, 5 mM K₃Fe(CN)₆, 5 mM K₄Fe(CN)₆ and 1 mg ml⁻¹ X-gal in PBS). Embryos were embedded in paraffin, sectioned and counterstained with nuclear fast red.

Generation and characterization of human iPSCs. Human BJ foreskin fibroblasts (Stemgent) were reprogrammed using published protocols^{35,36}, with the following modifications. BJ fibroblasts were first infected with lentivirus vectors encoding both a puromycin resistance gene and doxycycline-inducible shRNA targeting either GFP (negative control, target sequence: 5'-GCAAGCTGACCCCTGAAGTTCAT-3') or *MBNL1* shRNA (target sequence: 5'-GCCTGCTTTGATT CATTGAAA-3'). Lentiviral vector preparations and infections were performed as described³⁵. After selection with 1 $\mu\text{g ml}^{-1}$ puromycin, shRNA-encoding BJ fibroblasts were infected with a second lentivirus vector (obtained from Addgene) co-expressing both the mouse retrovirus receptor *mSlc7a1* and the blasticidin resistance gene³⁷. During transient selection with 5 $\mu\text{g ml}^{-1}$ blasticidin, puromycin was reduced to 0.5 $\mu\text{g ml}^{-1}$ and maintained at this concentration for 6 days after infection with retroviral reprogramming vectors.

pMXs-based retrovirus vectors encoding the four reprogramming factors hOCT4, hSOX2, hKLF4, hCMYC (OSKM)³⁷, were obtained from Addgene and packaged exactly as described³⁵. Puromycin/blasticidin-resistant BJ fibroblasts were infected in triplicate, using three separate preparations of retrovirus vectors. shRNA expression was induced by treatment with 2 $\mu\text{g ml}^{-1}$ doxycycline, which was initiated contemporaneously with retrovirus vector infection; control cells were treated with vehicle only.

Six days after retrovirus infection, BJ fibroblasts were seeded on a monolayer of feeder cells. Embryonic day 12.5 fibroblasts from Tg(DR4)1Jae/J mice (Jackson Laboratory) were seeded on collagen-coated 6-well plates at a density of 3×10^5 cells per well as described³⁵; retrovirus-infected BJ fibroblasts were seeded at a density of 2×10^4 cells per well. At day 28 of reprogramming, quantification (by whole-well morphological examination and by TRA1-60 immunostaining) of human iPSC colonies was performed by investigators who were blinded as to the experimental conditions. To count TRA1-60-positive colonies, the plates were imaged (for both TRA1-60-immunostained and DAPI channels) using an IN Cell Analyzer 2000 (GE Healthcare) with a $\times 4$ objective. For each well, 64 non-overlapping fields were captured and images were analysed using the Columbus System (PerkinElmer). Knockdown of MBNL1 resulted in an approximate twofold increase in TRA1-60 immunostaining colonies over the control knockdown with GFP-targeting shRNA (data not shown). Additional OSKM retrovirus-infected BJ fibroblasts were seeded in parallel; individual colonies from doxycycline-treated plates were manually isolated 4 weeks after infection, and seeded on feeders in collagen-coated 24-well plates³⁵. Cells from these colonies were expanded, and subsequently characterized (Supplementary Fig. 15) as described³⁸.

Clonal analysis by RNA-seq during reprogramming. In a single-cell assay, secondary MEFs were plated in individual wells of a 96-well plate, OKSM factors were induced by doxycycline treatment and the clonal derivatives were cultured for 21 days. Removal of doxycycline at day 21 revealed that approximately 50% of clones produced abundant alkaline-phosphatase-positive colonies (transgene-independent

clones), whereas the rest yielded few or no colonies (transgene-dependent clones). RNA-seq analysis was performed for three transgene-independent and five transgene-dependent clones at day 21 after doxycycline induction (Supplementary Table 1).

Using RNA-seq derived-PSI values (see Supplementary Methods), the inclusion levels of mouse ES-cell-differential cassette alternative exons were quantified for each of these clones. Fifty-one ES-cell-differential alternative splicing events with sufficient read coverage in all samples and with a ≥ 25 PSI difference between iPSCs and MEFs were compared between the two types of clones (Supplementary Fig. 14 and Supplementary Table 5).

RNA-seq data and analysis. We used RNA-seq data from 36 and 32 different human and mouse samples, respectively. Details and sample sources are provided in Supplementary Table 1. The samples comprise, for human: 5 ES-cell lines (3 different cell lines and 2 replicates), 2 iPSC lines, 7 non-ES-cell lines, and 22 adult tissues (18 different tissues and 4 replicates); for mouse: 6 ES-cell lines, 2 iPSC lines, 8 non-ES-cell lines (5 different cell lines and 3 replicates) and 16 adult tissues (10 different types and 6 replicates). Details of RNA-seq analysis is available in Supplementary Information.

To identify alternative exons differentially regulated in ES cells, we first calculated a single averaged PSI value for tissues of similar origin (see Supplementary Table 1). Only events with enough coverage in at least two ES-cell samples and three distinct tissue types were considered. 'ES-cell-differential alternative splicing events' were defined as those with a mean PSI difference of ≥ 25 between ES cells and differentiated tissues. To account for alternative splicing events potentially related to cell proliferation, we also required a mean PSI difference of ≥ 25 between ES-cell lines and non-ES-cell lines, when the event had sufficient coverage in at least one cell line sample. The set of background alternative splicing events used throughout the study are alternatively spliced exons (defined here as exons with PSI values of $< 95\%$ and $> 5\%$ in at least one sample) that meet the same expression requirement (that is, in ≥ 2 ES cells and ≥ 3 differentiated tissue types) and that show an average difference in PSI level of $< 5\%$ between ES cells and differentiated tissue samples, and between the ES-cell and non-ES-cell lines.

Analyses of splicing factor expression. A total of 221 human and 214 mouse genes were selected for analysis based on literature mining for previously described splicing functions, 'splicing'- and/or 'spliceosome'-associated Gene Ontology (GO) terms, and/or the presence of a PFAM-annotated RNA-binding domain (Supplementary Table 4). To calculate the mRNA expression values for each sample, we used corrected (for mappability) reads per kilobase pair and million mapped reads values (cRPKM) of the 'stable' (as defined by BioMart) Ensembl transcript for each gene, as previously described³⁹.

Splicing factor genes were ranked according to the relative extent of their differential expression (as determined by cRPKM values) in ES cells and iPSCs versus non-ES-cell lines and tissues by comparing summed ranks of each gene in all ES cell/iPSCs across the full range of samples. On the basis of this approach, human *MBNL1* and *MBNL2* showed the first and second lowest overall rank in ES cells/iPSCs, respectively, and mouse *Mbnl1* and *Mbnl2* showed the first and third lowest overall rank in ES cells/iPSCs, respectively.

To assess the statistical significance of the differential expression of individual splicing factor genes, we compared their cRPKM values in ES cells/iPSCs to the cRPKM values in all other cell lines and differentiated tissues using a Wilcoxon rank-sum test after quantile normalization. Splicing factors with Bonferroni-corrected *P*-values < 0.05 were considered significantly differentially expressed (Supplementary Table 4).

Splicing code analyses. The feature vectors for each species were produced by extracting sequence-based features from alternatively spliced exons, their adjacent constitutive exons, and 300 nucleotides of flanking intronic sequence. The features used were a subset of those defined previously⁴⁰, with the following differences: (1) all sequence length features are now in the log domain; (2) owing to a lack of comprehensive transcript libraries and the corresponding uncertainty about downstream consequences of frame shifts, premature termination codon features were excluded; and (3) conservation scores and conservation-weighted motifs were excluded from the feature set. In addition, related features (that is, consensus recognition sequences for a given splicing factor inferred by different methods) were combined and included as independent features.

To identify features strongly associated with ES-cell-differential exon inclusion or exclusion, we compared 172 ES-cell-differential exons and 908 background exons for human, along with 102 ES-cell-differential exons and 811 background exons for mouse. Associations between features and ES-cell-differential splicing were detected using Pearson correlation. For each feature, we computed the correlation between its value and the difference in average PSI values in ES cells and non-ES cells, across exons. To obtain more accurate correlation values, we considered two scenarios: (1) a positive scenario in which the differences in average PSI values in ES cells and non-ES cells are larger than 25%; and (2) negative

scenario in which the differences in average PSI values in ES cells and non-ES cells are smaller than -25% .

CLIP-seq analysis. We used recently described *Mbnl1* CLIP-seq data from C2C12 cells⁴¹. To estimate the fractions of ES-cell-differential and background alternative splicing events that are associated with MBNL1 binding, we asked whether CLIP binding clusters overlap the alternative exon and/or flanking intron sequences of each event. CLIP binding clusters were defined as previously described⁴¹. In short, CLIP-seq tags were trimmed of adapters and then collapsed to remove redundant sequences. These tags were mapped to genome and a database of splice junctions using Bowtie. To identify CLIP clusters lying within genic regions, gene boundaries were first defined using RefSeq, Ensembl and UCSC tables. For each window of 30 nucleotides covered by at least one CLIP-seq tag, a test was performed to assess whether the tag density in the window exceeded that which is predicted by a simple Poisson model which accounts for gene expression and pre-mRNA length. An alternative splicing event was considered to have an overlapping MBNL1 binding cluster if the mid-point of the cluster is located within the alternative exon, within 300 nucleotides of the 5' or 3' ends of upstream or downstream flanking introns, and/or within 30 nucleotides within the 3' end of C1 exon or the 5' end of C2 exon. Only alternative splicing events that had significant read coverage (see above) in at least one of two C2C12 samples used were analysed. In total, 57 ES-cell-differential alternative splicing events and 601 background alternative splicing events were compared.

To generate an RNA regulatory map⁴² highlighting MBNL1 binding sites in relation to ES-cell-differential alternative splicing events with either higher (ES-cell-included) or lower (ES-cell-excluded) exon inclusion levels in ES cells versus other cell lines and tissues, we applied the following procedure: for each nucleotide position from the regions described above and from three sets of alternative splicing events (ES-cell-included, ES-cell-excluded and background), we counted the average number of MBNL1 CLIP-seq tags. To minimize the impact of outliers with extreme read density, we limited the maximum count per event to an average of ten reads per position within each region. To normalize the length of the alternative splicing exon, we divided each exon into 100 bins and uniquely assigned each position to the integer of $100 \times \text{position/exon_length}$, with a relative weight inversely related to the length of the alternative splicing exon. To draw the map, we used sliding windows of 30 nucleotides for the intronic regions and 25 nucleotides for the length-corrected exons (total of four windows shown).

Evolutionary conservation of ES-cell-differential events. We analysed three different aspects of conservation of the human and mouse ES-cell-differential alternative exons⁴³. To determine whether the alternative exon is conserved at the genomic level, we performed a lift-over of the exon coordinates using Galaxy (<https://main.g2.bx.psu.edu/>). Exons with a unique lift-over hit in the other species, and with AG (splicing acceptor) and/or GT (splicing donor) sites were considered to be genome-conserved in the other species. In addition, if the orthologous exon has a PSI of $< 95\%$ and $> 5\%$ in at least one sample from each species, alternative splicing of the exon was defined as conserved. Finally, to assess whether ES-cell-differential regulation is conserved, we applied two criteria: (1) the exons are independently detected as ES-cell-differential in human and mouse using the criteria as described above (total = 25 alternative splicing events); and (2) the orthologous exons must meet minimal read coverage requirements (also as described above) to afford direct comparison.

Analyses of function and protein domain enrichment. To investigate whether ES-cell-differential events are significantly enriched in genes with specific functional associations and/or protein domains, we used the online tool DAVID (<http://david.abcc.ncifcrf.gov/>)^{44,45} (with annotations from levels 3, 4 and 5 in the GO hierarchy), KEGG pathways and InterPro domains. As background, we used the genes with at least one alternative splicing event that met the minimal expression criteria described above (that is, detection in ≥ 2 ES cells and ≥ 3 differentiated cell/tissue types). The main clusters of functionally related genes enriched in both human and mouse (as well as among the conserved) ES-cell-differential events (Supplementary Table 3) are associated with: actin cytoskeleton, plasma membrane (including cell junctions) and protein kinase-associated terms.

- Gabut, M. *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147**, 132–146 (2011).
- Polo, J. M. & Hochedlinger, K. When fibroblasts MET iPSCs. *Cell Stem Cell* **7**, 5–6 (2010).
- Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
- Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11**, 769–782 (2012).
- Hotta, A. *et al.* EOS lentiviral vector selection system for human induced pluripotent stem cells. *Nature Protocols* **4**, 1828–1844 (2009).

36. Hotta, A. *et al.* Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. *Nature Methods* **6**, 370–376 (2009).
37. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
38. Cheung, A. Y. *et al.* Isolation of MECP2-null Rett Syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. *Hum. Mol. Genet.* **20**, 2103–2115 (2011).
39. Labbé, R. M. *et al.* A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**, 1734–1745 (2012).
40. Xiong, H. Y., Barash, Y. & Frey, B. J. Bayesian prediction of tissue-regulated splicing using RNA sequence and cellular context. *Bioinformatics* **27**, 2554–2562 (2011).
41. Wang, E. T. *et al.* Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150**, 710–724 (2012).
42. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
43. Irimia, M., Rukov, J. L., Roy, S. W., Vinther, J. & Garcia-Fernandez, J. Quantitative regulation of alternative splicing in evolution and development. *Bioessays* **31**, 40–50 (2009).
44. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
45. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).

The bromodomain protein Brd4 insulates chromatin from DNA damage signalling

Scott R. Floyd^{1,2}, Michael E. Pacold^{1,3,4}, Qiuying Huang¹, Scott M. Clarke¹, Fred C. Lam¹, Ian G. Cannell¹, Bryan D. Bryson¹, Jonathan Rameseder¹, Michael J. Lee¹, Emily J. Blake¹, Anna Fydrych¹, Richard Ho¹, Benjamin A. Greenberger¹, Grace C. Chen¹, Amanda Maffa¹, Amanda M. Del Rosario¹, David E. Root⁵, Anne E. Carpenter⁵, William C. Hahn^{5,6}, David M. Sabatini^{4,5}, Clark C. Chen^{6,7}, Forest M. White^{1,8}, James E. Bradner^{5,6} & Michael B. Yaffe^{1,5,7,8,9}

DNA damage activates a signalling network that blocks cell-cycle progression, recruits DNA repair factors and/or triggers senescence or programmed cell death¹. Alterations in chromatin structure are implicated in the initiation and propagation of the DNA damage response². Here we further investigate the role of chromatin structure in the DNA damage response by monitoring ionizing-radiation-induced signalling and response events with a high-content multiplex RNA-mediated interference screen of chromatin-modifying and -interacting genes. We discover that an isoform of Brd4, a bromodomain and extra-terminal (BET) family member, functions as an endogenous inhibitor of DNA damage response signalling by recruiting the condensin II chromatin remodelling complex to acetylated histones through bromodomain interactions. Loss of this isoform results in relaxed chromatin structure, rapid cell-cycle checkpoint recovery and enhanced survival after irradiation, whereas functional gain of this isoform compacted chromatin, attenuated DNA damage response signalling and enhanced radiation-induced lethality. These data implicate Brd4, previously known for its role in transcriptional control, as an insulator of chromatin that can modulate the signalling response to DNA damage.

Detection and repair of damaged DNA is integral for cell survival and accurate transmission of genetic information to progeny. Defects in the DNA damage response (DDR) contribute to oncogenesis and genomic instability in tumours^{3,4} and render tumour cells sensitive to DNA-damaging cancer therapy⁵. Early signalling events that trigger and transduce the DDR occur in the context of chromatin, and it is likely that modulation of chromatin structure plays a role in DDR signalling². Histone proteins are known targets of DDR post-translational modification^{2,6}, but a detailed understanding of the role of chromatin modulation in the DDR is lacking.

To explore the role of chromatin modulation in the DDR, we developed a high-throughput, high-content quantitative microscopy assay multiplexed for early and late DDR endpoints, and applied this to an RNA-mediated interference (RNAi) library focused on proteins that interact with and modify chromatin (see Methods)^{7,8}. For each time point, cells were co-stained with γ H2AX antibodies to measure early signalling events in the DDR, Hoechst 33342 to monitor cell-cycle progression and phospho-histone H3 (pHH3) to measure mitotic entry. At the latest time point, cleaved caspase-3 (CC3) was substituted for pHH3 to measure apoptotic cell death. The screening assay was validated with small molecule inhibitors of DDR signalling as well as RNAi directed against known components of the DDR pathway (Supplementary Figs 1–4).

The most pronounced increase in γ H2AX foci number, size and intensity after ionizing radiation was observed at 1 and 6 h after

knockdown of Brd4; this remained elevated at 24 h (Fig. 1a, b and Supplementary Fig. 4). Eight hairpins directed against Brd4 showed this effect, making off-target effects unlikely (Fig. 1a and Supplementary Fig. 4). Neither Brd4 knockdown in the absence of irradiation (Fig. 1b) nor knockdown of other bromodomain-containing proteins (Figs 1b and Supplementary Fig. 4) significantly altered γ H2AX. Increased ionizing-radiation-induced γ H2AX after Brd4 loss was further confirmed using short interfering RNA (siRNA) oligonucleotides targeting additional independent Brd4 sequences (Fig. 1f and Supplementary Fig. 5).

Brd4 encodes three splice isoforms (A, B and C in Fig. 1c). Each isoform contains two amino (N)-terminal bromodomains (BD1 and BD2) that bind acetylated lysine, and an extra-terminal (ET) domain recently reported to interact with several chromatin-binding proteins⁹. The A isoform contains a carboxy (C)-terminal domain (CTD) that functions as a transcriptional co-activator with the pTEFb complex^{10,11}. This region is notably absent in the B and C isoforms, and in the B isoform it is replaced with a divergent short 75 amino-acid segment. All three Brd4 isoforms are expressed in U2OS cells, and the short hairpin RNAs (shRNAs) used in our screen targeted all three isoforms (Supplementary Table 1). We confirmed that a single distinct siRNA that was active against all Brd4 isoforms replicated the Brd4 loss-of-function phenotype of elevated ionizing-radiation-induced γ H2AX (Supplementary Fig. 5).

To establish the relative effects of the isoforms on the DDR, we performed gain-of-function experiments. Overexpression of Brd4 isoform B most potently suppressed ionizing-radiation-induced γ H2AX foci (Fig. 1d). We designed isoform-specific siRNAs to reduce expression of isoform A or B messenger RNA (mRNA) (Fig. 1e) and protein (Supplementary Fig. 5) selectively; selective targeting of isoform C was not technically possible owing to complete coding sequence overlap with isoforms A and B. We observed that selective depletion of Brd4 isoform B, but not isoform A, increased H2AX phosphorylation over a wide range of ionizing radiation doses (Fig. 1f).

To investigate whether elevated γ H2AX levels observed in Brd4-deficient cells resulted from increased production of ionizing-radiation-induced DNA double-strand breaks or from faulty double-strand break repair, we used pulsed-field gel electrophoresis to quantify double-strand breaks in control and Brd4 knockdown cells. As shown in Fig. 2a, Brd4 knockdown had minimal effects on the generation and repair kinetics of double-strand breaks. These observations, together with our finding that individual γ H2AX foci were larger and more intense in irradiated Brd4 knockdown cells (Fig. 1b, Supplementary Fig. 4 and Supplementary Tables 1 and 2), indicate that there is enhanced signalling from damaged DNA in the absence of Brd4, rather than an increase in the amount of damage or repair deficiency.

¹Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Department of Radiation Oncology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. ³Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ⁴Whitehead Institute, Cambridge, Massachusetts 02139, USA. ⁵Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ⁶Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ⁷Department of Surgery, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. ⁸Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁹Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

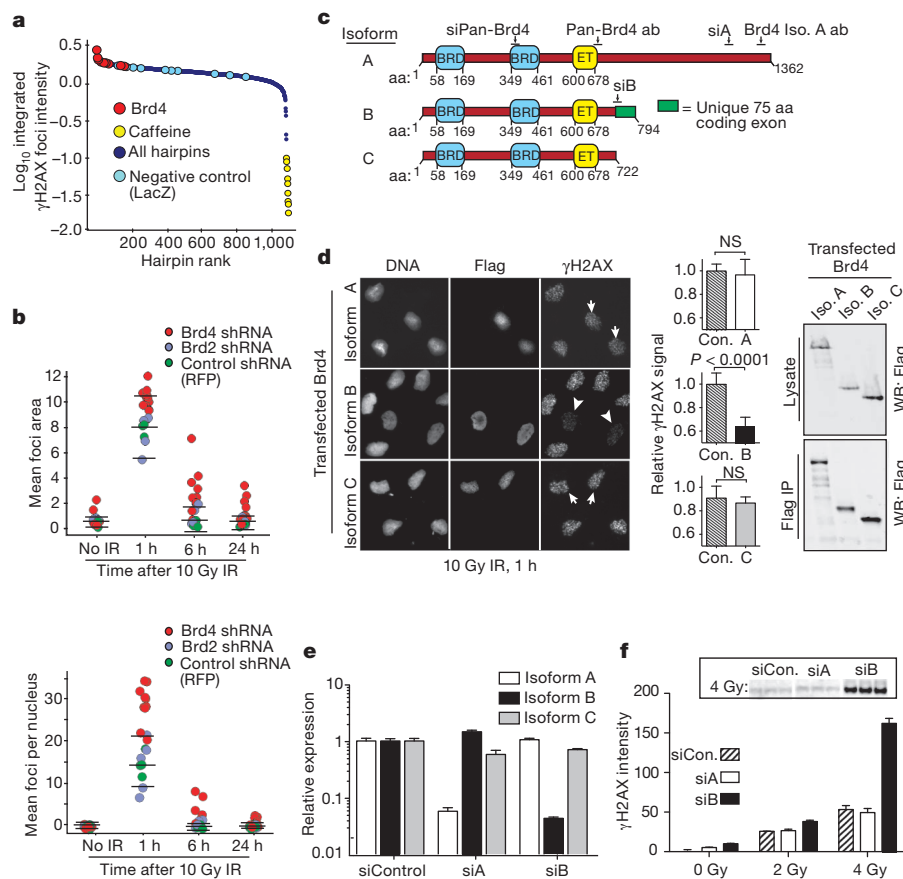


Figure 1 | Brd4 isoform B suppresses H2AX phosphorylation after ionizing radiation.

a, Rank of hairpins from shRNA screen ordered by integrated γ H2AX foci intensity at 1 h after 10 Gy ionizing radiation (details of screening assay in Supplementary Figs 1–4). **b**, γ H2AX foci size (upper panel) and mean γ H2AX foci per nucleus (lower panel) after 10 Gy ionizing radiation (IR) from cells expressing indicated shRNAs (bars show mean and two standard deviations of control values). RFP, red fluorescent protein. **c**, Domain structure of Brd4 isoforms showing conserved tandem bromodomains (BRD), extra-terminal (ET) domain, siRNA and antibody target sequences, and unique isoform B exon. aa, amino acid; Iso., isoform. **d**, H2AX phosphorylation in cells expressing Flag-tagged Brd4 isoform B (arrowheads) or A and C (arrows) at 1 h after 10 Gy IR. Left: representative images. Middle: quantification of 10 fields from two independent experiments with mean γ H2AX signal normalized to untransfected cells. Right: immunoblot of isoform expression levels in whole-cell lysates and anti-Flag immunoprecipitates. Con., control. **e**, Isoform-specific Brd4 knockdown in cells transfected with the indicated siRNA and analysed by quantitative real-time PCR with reverse transcription ($n = 3$). **f**, H2AX phosphorylation levels 1 h after indicated ionizing radiation exposure in cells transfected with isoform-specific siRNA ($n = 3$). Inset shows representative immunoblot for triplicate samples. Data are from U2OS cells. Error bars, s.e.m.; P values were determined using Student's t -test in this and all subsequent figures unless otherwise indicated.

Changes in overall chromatin structure can affect H2AX phosphorylation, probably by controlling the accessibility of signalling molecules to DNA damage sites^{12,13}. Interestingly, γ H2AX foci form more readily in 'open' areas of euchromatin¹⁴, histone acetylation has been linked to the 'open' chromatin state and histone deacetylase inhibitors are known to increase H2AX phosphorylation¹⁵. We speculated that a bromodomain protein could influence H2AX phosphorylation through interaction with acetylated histones and effects on global

chromatin structure, and therefore performed micrococcal nuclease susceptibility experiments. Knockdown of Brd4 isoform B increased digestion by micrococcal nuclease, indicating a more 'open' overall chromatin structure, whereas knockdown of isoform A had minimal effects (Fig. 2b). Furthermore, we observed that cells transfected with Brd4 isoform B showed a distinct nuclear 4',6-diamidino-2-phenylindole (DAPI) staining pattern, indicating a change in chromatin structure (Fig. 2c). As shown in Fig. 2d, e, quantification of the

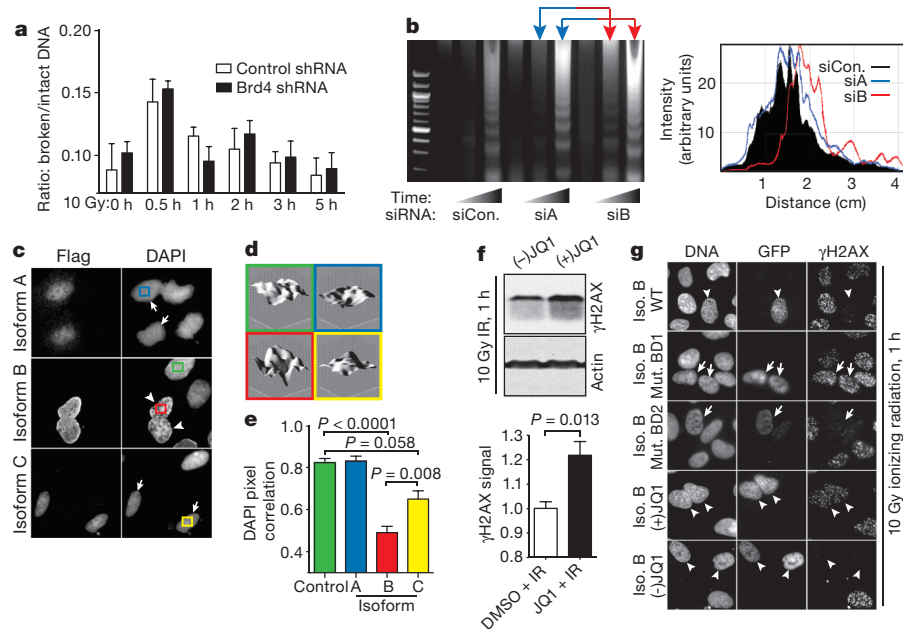


Figure 2 | Brd4 isoform B limits H2AX phosphorylation through bromodomain-acetyl lysine-mediated effects on chromatin structure.

a, Pulsed-field electrophoresis analysis of DNA from stable cell lines expressing indicated shRNA after 10 Gy IR ($n = 3$). **b**, Left: micrococcal nuclease assay of control or Brd4 knockdown cells. Right: line traces of representative gel lanes. **c**, Chromatin structure from cells expressing Flag-tagged Brd4 isoform B (arrowheads) or A and C (arrows) shown by DAPI staining. **d**, Three-dimensional representation of nuclear DAPI staining intensity from cells in **c** as indicated by coloured frames. **e**, DAPI pixel correlation from Brd4 isoform A, B, C and untransfected control cells ($n = 3$). **f**, Immunoblots (top) and quantification (bottom) of H2AX phosphorylation after 250 nM DMSO, or active (+) and inactive (–) JQ1 at 1 h after 10 Gy ionizing radiation ($n = 3$). **g**, γ H2AX signal 1 h after 10 Gy IR in cells expressing green fluorescent protein (GFP)–wild-type Brd4 isoform B (arrowheads), isoform B with mutations that abrogate acetyl lysine binding of bromodomain 1 (BD1) or 2 (BD2) (arrows), or wild-type Brd4 isoform B in the presence of 250 nM (–) JQ1 (inactive) or (+) JQ1 as indicated.

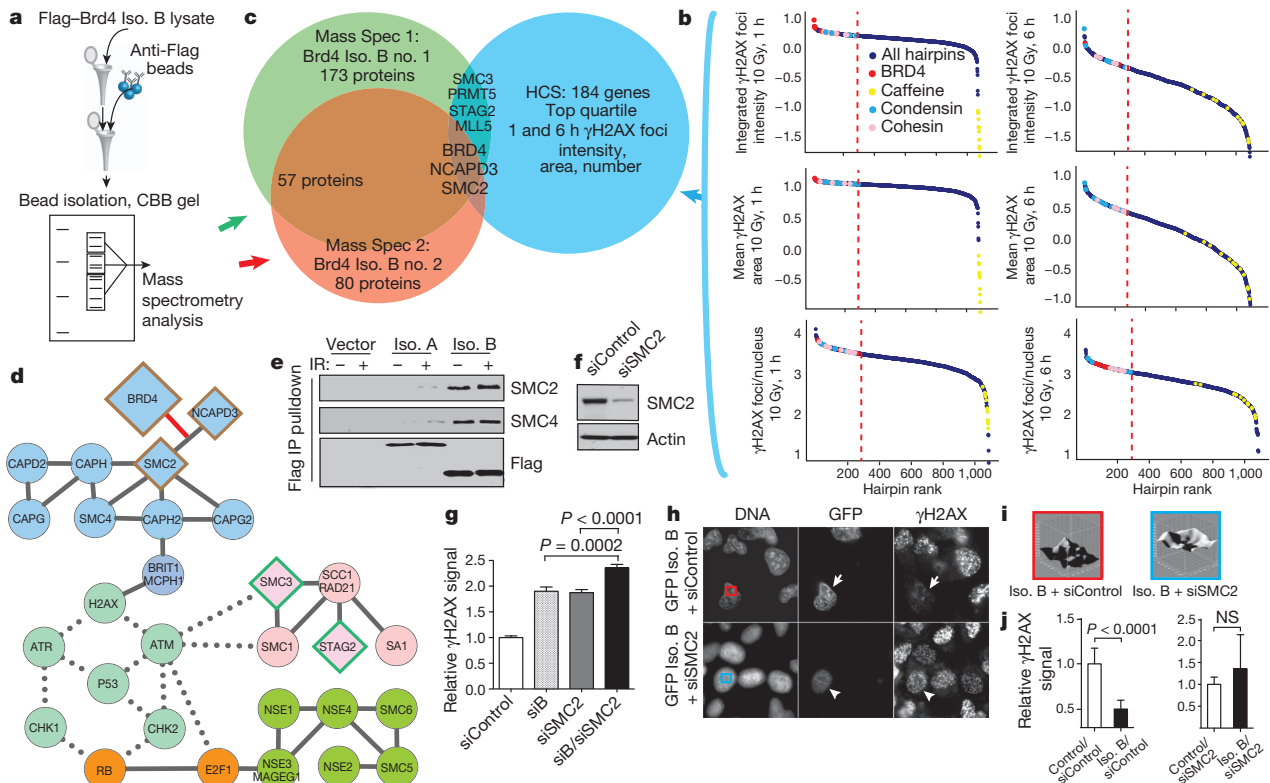


Figure 3 | Brd4 isoform B interaction with the condensin complex affects H2AX phosphorylation. **a**, Mass spectrometry identification of co-immunoprecipitated proteins from Flag-tagged Brd4 isoform-B-expressing cells. **b**, Identification of candidate Brd4 interactors by ranking chromatin modifier shRNAs from screen for elevated H2AX foci intensity, area and number at 1 and 6 h after 10 Gy IR. Dashed red lines indicate top quartile. **c**, Intersection of two independent mass-spectrometry experiments (**a**) with the top quartile of candidates in **b**. Overlapping set includes Brd4, SMC2 and NCAPD3. **d**, Network representation of SMC proteins and relation to DNA damage signalling with protein-protein and kinase-substrate interactions collated from the literature. Protein-protein and kinase-substrate interactions shown by solid and dotted lines, respectively. Colours indicate condensin complex (blue), cohesin complex (pink), other SMC protein complexes (green), cell-cycle regulators (orange) and DNA damage signalling machinery (mint). Diamonds show mass spectrometry and high-content screening hits from **a** and **b**. **e**, Validation of isoform-B-condensin interaction with blotting immunoprecipitates from cells transfected with indicated Flag-tagged constructs. **f**, Immunoblot verification of SMC2 knockdown from cells transfected with SMC2 siRNA. **g**, Nuclear γ H2AX signal from cells transfected with indicated combinations of control DNA, Brd4 isoform B and/or SMC2 siRNA. Data were quantified from ten fields of two independent experiments normalized to control cells. **h**, H2AX phosphorylation 1 h after 10 Gy IR in cells simultaneously expressing isoform B and control (arrows) or SMC2 siRNA (arrowheads). **i**, Chromatin staining pattern in cells simultaneously expressing isoform B and control (red frame) or SMC2 (blue frame) siRNA. **j**, Mean nuclear γ H2AX signal in GFP-isoform-B-expressing cells with or without SMC2 knockdown. Data are from ten fields of two independent experiments, as in **h**, normalized to control untransfected cells.

nuclear staining texture showed a more heterogeneous DAPI intensity pattern, and significantly lower pixel-to-pixel correlation of DAPI staining in cells overexpressing isoform B, indicative of isoform-B-mediated alterations in global chromatin structure. Expression of isoform A had no effect on DAPI staining, whereas overexpression of isoform C had smaller effects than those observed with isoform B.

Our finding that Brd4 isoform B expression affects global chromatin structure and attenuates H2AX phosphorylation in response to DNA damage led us to investigate the subcellular localization of isoform B in response to ionizing radiation. Immunofluorescence experiments showed that ionizing radiation did not grossly alter Brd4 isoform B nuclear localization, which tightly mirrored DNA patterns shown by DAPI staining (Supplementary Fig. 6a). Interestingly, subcellular fractionation of U2OS cells and extraction of chromatin-bound proteins demonstrated that irradiation caused enhanced isoform B association with the high salt-extractable chromatin fraction (Supplementary Fig. 6b, c), indicating increased association of isoform B with chromatin after DNA damage.

Bromodomains recognize epigenetic marks on chromatin by binding to acetyl-lysine¹⁶. We therefore tested the contribution of Brd4 bromodomain interactions to alterations in γ H2AX phosphorylation using JQ1, a small molecule inhibitor of BET bromodomains¹⁷. Only the active enantiomer of JQ1 caused increased H2AX phosphorylation

after irradiation in U2OS cells (Fig. 2f), similar to the effects observed after Brd4 isoform-B-specific knockdown. Furthermore, JQ1 treatment or Brd4 isoform B knockdown did not significantly alter total histone levels or levels of histone acetylation (Supplementary Figs 7 and 8). Interestingly, overexpression of Brd4 isoform B led to alteration in the nuclear staining pattern of acetyl-lysine, closely mirroring the DAPI staining pattern induced by expression of isoform B (Supplementary Fig. 7b).

The concentration of JQ1 that we used (250 nM) is consistent with the reported *in vitro* half-maximum inhibitory concentration for Brd4 bromodomains 1 (BD1, 77 nM) and 2 (BD2, 33 nM)¹⁷. To evaluate directly the role of each bromodomain in isoform B, we performed gain-of-function experiments using wild-type Brd4 in the absence or presence of JQ1, or constructs harbouring mutations that abrogate acetyl lysine binding by BD1 or BD2. Mutations in BD1, or addition of the active enantiomer of JQ1, potentially reversed the γ H2AX-suppressive effects of isoform B expression (Fig. 2g). Notably, mutations that abrogate BD1 binding to acetyl-lysine also rescued the ionizing-radiation-induced cell death phenotype observed with Brd4 isoform B gain-of-function (see below), implicating BD1 in the mechanism of DNA damage inhibition (cf. Fig. 4b).

To probe further the role of lysine acetylation on γ H2AX-Brd4 effects, we examined the combined effects of histone deacetylase

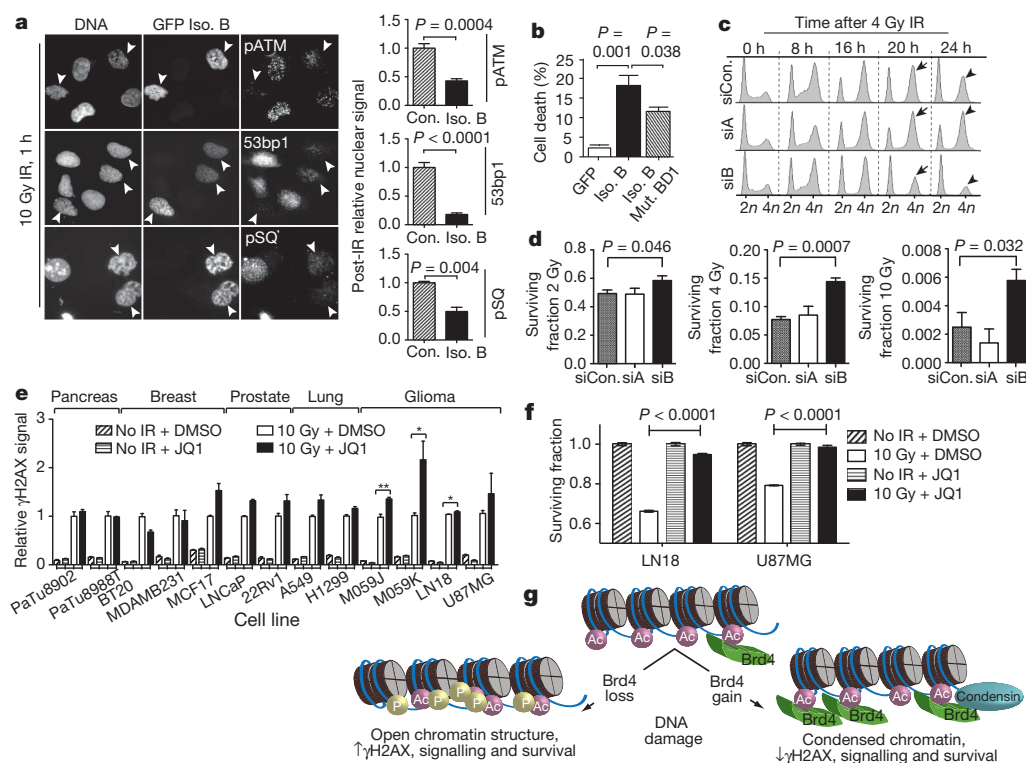


Figure 4 | Brd4 isoform B affects ionizing-radiation-induced cell-cycle checkpoints and survival.

a, Loss of DNA damage signalling in cells expressing Brd4 isoform B. Left: representative images stained for indicated DDR proteins 1 h after 10 Gy IR. Arrowheads indicate isoform-B-expressing cells. Right: quantification of ten representative fields from two independent experiments normalized to untransfected cells. **b**, Cell death 24 h after 10 Gy IR in cells expressing wild-type or bromodomain-1-mutant isoform B (Isoform B mut. BD1) scored for cleaved caspase 3 by flow cytometry ($n = 3$). **c**, Ionizing-radiation-induced cell-cycle arrest and recovery in Brd4 isoform knockdown cells assayed by propidium iodide staining and flow cytometry. **d**, Cell survival after irradiation in Brd4 isoform knockdown cells measured by colony formation. **e**, JQ1 effect on γ H2AX in several human cancer cell types commonly treated with radiotherapy. No IR, no ionizing radiation. **f**, Radiation survival effects of JQ1 in glioma cell lines measured at 72 h by CellTiterGlo ($n = 3$). **g**, Model for Brd4 effects on DNA damage signalling.

inhibitors and Brd4 knockdown. We found that when Brd4 isoform B knockdown was combined with exposure to 50 nM LBH589, an inhibitor of histone deacetylases 1–3 and 6 (ref. 18), H2AX phosphorylation was enhanced to a greater extent than with either treatment alone (Supplementary Fig. 9). This effect could be observed even in unirradiated cells, although the total amount of H2AX phosphorylation remained lower than that seen in irradiated cells. Taken together, these findings indicate that Brd4 isoform B binding to acetylated regions of chromatin alters chromatin structure and limits H2AX phosphorylation.

Brd4 also has a defined role in transcriptional modulation, largely through interactions of isoform A with the pTEFb transcriptional complex^{10,11}. To investigate the contribution of Brd4-driven transcriptional changes to the suppression of DNA damage signalling, we profiled mRNA expression patterns of cells stably expressing control or Brd4 shRNAs. Only one DDR-associated transcript, CHEK2, showed a differential expression change of twofold or more (Supplementary Fig. 10a). Importantly, transient Brd4 knockdowns with siRNA, or short-term inhibition with JQ1, both of which increased γ H2AX foci formation after irradiation (Supplementary Fig. 5a and Fig. 2f), caused no change in CHEK2 mRNA levels (Supplementary Fig. 10b, c), and neither long-term nor short-term Brd4 knockdown affected the protein levels of several DDR molecules, including Chk2 (Supplementary Fig. 10d). Moreover, the suppression of DDR signalling by Brd4 isoform B overexpression was insensitive to transcription and translation inhibition with α -amanitin and cycloheximide, respectively (Supplementary Fig. 11).

As interactions between Brd4 and other protein complexes involved in modulating chromatin structure were probably responsible for the DDR effects we observed, we identified proteins co-immunoprecipitated with isoform B after DNA damage using mass spectrometry (Fig. 3a and Supplementary Fig. 12). From two independent experiments, we obtained a common set of 57 interacting proteins (Supplementary Tables 3 and 4). Because the DDR-relevant Brd4-binding proteins presumably function in the same pathway as Brd4, we reasoned that loss of these proteins should show a phenotype similar to Brd4 loss-of-function. We therefore used our existing high-content screening data to create a

list of the top quartile of genes ranked by increased γ H2AX foci intensity, number and size at 1 and 6 h after irradiation (Fig. 3b). The overlap of this list with the list of isoform-B-interacting proteins showed two members of the condensin II complex, SMC2 and CAPD3 (Fig. 3c, d). This finding was intriguing as the condensin II complex has a known role in chromatin compaction in both mitotic and interphase cells, and has been linked to DNA damage repair¹⁹. We performed immunoprecipitation experiments after DNA damage, and found that the SMC2 and SMC4 components of the condensin II complex co-immunoprecipitated with Brd4 isoform B, whereas Brd4 isoform A had minimal co-association (Fig. 3e). To verify the role of this interaction on the γ H2AX effects we observed, we performed combined isoform B and SMC2 knockdown and assayed H2AX phosphorylation 24 h after siRNA transfection, when knockdown of each protein is sub-maximal. We found that H2AX phosphorylation was enhanced with combined knockdown over knockdown of either protein alone (Fig. 3f, g). Furthermore, in cells overexpressing isoform B, SMC2 knockdown could abrogate the suppressive effects of Brd4 on γ H2AX, demonstrating a functional interaction between isoform B and the condensin II complex in modulating γ H2AX (Fig. 3h, j). Finally, we noted that the effects of isoform B on the DAPI staining pattern of chromatin were abrogated by co-transfection of SMC2 siRNA, indicating that the Brd4-condensin II interaction is involved in chromatin structure alterations (Fig. 3i).

We next investigated isoform B effects on other components of the DDR. We found that isoform B gain-of-function inhibited ionizing-radiation-induced foci formation of several other known DDR signalling components including 53BP1, phosphorylated ATM and several DDR signalling molecules containing the phospho-SQ DDR kinase substrate motif (Fig. 4a). In addition, overexpression of isoform B resulted in increased cell death after irradiation, an effect that was significantly diminished by mutation of BD1 (Fig. 4b). The cell death observed in Brd4 isoform B overexpressing cells seems to result from mitotic catastrophe, consistent with a loss of DDR signalling that results in failed cell-cycle arrest (Supplementary Fig. 13). We also investigated the effect of isoform B knockdown on DDR-induced cell-cycle arrest and survival. Interestingly, isoform B loss-of-function

allowed increased cell survival with more rapid and efficient recovery from cell-cycle arrest after irradiation, complementing the inverse findings observed with isoform B gain-of-function (Fig. 4c, d).

Given the effects of Brd4 isoform B on ionizing-radiation-induced DDR signalling and survival, we considered that isoform B might have a role in tumour responses to irradiation. We screened a panel of established cell lines from several human tumour types commonly treated with radiotherapy for γ H2AX effects using the JQ1 inhibitor. Several cell types showed increased ionizing-radiation-induced H2AX phosphorylation with JQ1 treatment, including breast, prostate and particularly glioma cancer cell lines (Fig. 4e). Just as we had observed with U2OS cells, irradiation had the expected killing effect on dimethylsulphoxide (DMSO)-treated glioma cells; however, this killing effect was markedly reduced in JQ1-treated glioma cells, consistent with our finding of increased DDR signalling and radioresistance with decreased Brd4 function (Fig. 4f). Conversely, overexpression of Brd4 isoform B in glioma cells inhibited H2AX phosphorylation, consistent with decreased DDR signalling upon Brd4 gain-of-function (Supplementary Fig. 14).

We conclude that structural alterations in chromatin mediated by Brd4 acetyl lysine binding function to attenuate the DNA damage signalling response to ionizing radiation. These effects on DDR signalling are consistent with the induction of a chromatin structure that is inhibitory to the formation of γ H2AX in the case of higher levels of Brd4 isoform B expression, or a more 'open' chromatin structure that facilitates γ H2AX foci formation when Brd4 expression is reduced, or after pharmacological inhibition of bromodomain binding (shown schematically in Fig. 4g).

Our data indicate that Brd4 affects DDR signalling through mechanisms distinct from known transcriptional interactions with the P-TEFb transcriptional complex. The relevant Brd4 isoform that modulates the DDR, isoform B, lacks the pTEFb-interacting region. In addition, chemical inhibition of transcription/translation had no effect on the ability of Brd4 to suppress DDR-induced γ H2AX. This finding is in line with the recent identification of other chromatin-interacting proteins such as KAP-1 and Brg1 that have roles in DNA damage signalling that do not seem to arise directly from the transcriptional activity that these molecules also possess^{13,20}. Rather, the enhancement of several parameters of γ H2AX foci after Brd4 knockdown, including their size and intensity, in addition to their number, point to a role for Brd4 in limiting the propagation of DDR signalling after ionizing radiation. This effect seems to involve the recruitment of a chromatin-condensing complex to sites of acetylation, a new role for Brd4. In agreement with this, overexpression of Brd4, even in the absence of damage, resulted in alterations of chromatin structure and nuclear acetylation patterns, consistent with a model of Brd4 isoform B binding to and occluding acetyl-lysine sites on chromatin and recruiting chromatin compaction machinery. These findings implicate bromodomain-mediated interactions in modulating specific chromatin structures that inhibit the propagation of DDR signalling in chromatin^{12,15}, and indicate that Brd4 isoform B alters the threshold response of γ H2AX to DNA damage.

METHODS SUMMARY

Image-based high-content screening was performed in 384-well plate format using an arrayed lentiviral shRNA library from The RNAi Consortium. Screen images were acquired with a Cellomics microscope (Thermo Scientific) and quantified using CellProfiler software. siRNAs and antibodies were from commercial sources. We used Affymetrix U133 Plus 2.0 arrays for expression profiling. Mass spectrometry data from Brd4 immunoprecipitates after SDS-PAGE was acquired with an Orbitrap XL instrument (Thermo Scientific), and data analysed with Mascot software. Interactions for network analysis were hand-curated from primary literature using the keywords 'DNA damage', 'cell cycle checkpoint', 'chromatin structure', 'ATM/ATR', 'Chk1/Chk2' and 'SMC proteins'. Further details are provided in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 19 July 2011; accepted 3 April 2013.

Published online 2 June 2013.

1. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
2. Misteli, T. & Soutoglou, E. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nature Rev. Mol. Cell Biol.* **10**, 243–254 (2009).
3. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
4. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
5. Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
6. Polo, S. E. & Jackson, S. P. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev.* **25**, 409–433 (2011).
7. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
8. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
9. Rahman, S. *et al.* The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol. Cell Biol.* **31**, 2641–2652 (2011).
10. Yang, Z. *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* **19**, 535–545 (2005).
11. Jang, M. K. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19**, 523–534 (2005).
12. Murga, M. *et al.* Global chromatin compaction limits the strength of the DNA damage response. *J. Cell Biol.* **178**, 1101–1108 (2007).
13. Ziv, Y. *et al.* Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nature Cell Biol.* **8**, 870–876 (2006).
14. Cowell, I. G. *et al.* γ H2AX foci form preferentially in euchromatin after ionising radiation. *PLoS ONE* **2**, e1057 (2007).
15. Kim, J. A., Kruhlak, M., Dotiwala, F., Nussenzweig, A. & Haber, J. E. Heterochromatin is refractory to γ -H2AX modification in yeast and mammals. *J. Cell Biol.* **178**, 209–218 (2007).
16. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214–231 (2012).
17. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
18. Bradner, J. E. *et al.* Chemical phylogenetics of histone deacetylases. *Nature Chem. Biol.* **6**, 238–243 (2010).
19. Wu, N. & Yu, H. The SMC complexes in DNA damage response. *Cell Biosci.* **2**, 5 (2012).
20. Lee, H.-S., Park, J.-H., Kim, S.-J., Kwon, S.-J. & Kwon, J. A cooperative activation loop among SWI/SNF, γ -H2AX and H3 acetylation for DNA double-strand break repair. *EMBO J.* **29**, 1434–1445 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Le, T.R. Jones and M. Vokes for assistance with screening and image analysis. We thank C. Whittaker, S. Hoersch and M. Moran for computing and data analysis assistance; C. Reinhardt, C. Ellison and A. Gardino for manuscript editing; and P. Filippakopoulos and S. Knapp for discussions. This work was partially supported by the Koch Institute and Center for Environmental Health Sciences National Institutes of Health Core Grants P30-CA14051 and ES-002109; and by grants R01-ES15339, 1-U54-CA112967-04 and R21-NS063917; a SPARC grant to M.B.Y.; and a Holman Pathway Research Resident Seed Grant, American Society for Radiation Oncology Junior Faculty Career Research Training Award, Klarman Scholar, Koch Institute Clinical Investigator Award, and Burroughs Wellcome Career Award for Medical Scientists to S.R.F.

Author Contributions S.R.F. and M.B.Y. designed the study, supervised the experiments, analysed the data and wrote the manuscript. D.E.R., W.C.H. and D.M.S. were involved in the design and preparation of the lentiviral shRNA library. S.R.F., M.E.P. and E.B. performed the image-based high-content screen and initial analysis. A.E.C. aided in digital image analysis. S.R.F., Q.H., S.M.C., F.C.L., I.G.C., M.J.L., A.F., R.H., B.A.G., G.C.C. and A.M. performed biochemical, cell biological and molecular biological experiments. B.D.B., A.M.D. and F.M.W. performed mass spectrometry experiments and analysis. J.R. performed bioinformatics analysis. J.E.B. contributed JQ1 compounds and cell lines. S.R.F. and M.B.Y. designed and supervised the experiments. C.C.C., J.E.B. and F.M.W. contributed to the intellectual development of the study and technical writing of the manuscript. All authors contributed to editing the manuscript.

Author Information The expression profiling Affymetrix u133 plus dataset has been deposited in the NCBI Gene Expression Omnibus database under accession number GSE30700. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B.Y. (myaffe@mit.edu).

METHODS

Antibodies and stains. Mouse monoclonal antibodies against γ H2AX were from Upstate/Millipore (catalogue number 05636), Actin (Sigma, catalogue number A5441), phospho-ATM Serine 1981 (Rockland, catalogue number 200-301-400), Flag (Sigma, catalogue number F3165), ornithine decarboxylase (Abcam, catalogue number ab66067), RAD50 (GeneTex, catalogue number GTX70228), NBS1 (Abcam, catalogue number ab49958), MDC1 (Novus, catalogue number NB100-396) and Lamin (Millipore, catalogue number 05-714). Rabbit polyclonal and monoclonal antibodies against Brd4 were from Abcam (catalogue number Ab46199) and Pan-Brd4 from Sigma (catalogue number AV39076), 53BP1 (Novus, catalogue number NB100-304), CHEK2 (Cell Signaling Technology, catalogue number 2662), total H2AX (Abcam, catalogue number ab11175), phospho-SQ (Cell Signaling Technology, catalogue number 2851), MRE11 (Novus, catalogue number NB100-142), cleaved caspase 3 (Cell Signaling Technology, catalogue number 9664), SMC2 (Cell Signaling Technology, catalogue number 5329), SMC4 (Cell Signaling Technology, catalogue number 5547), phospho-histone H3 (Upstate/Millipore, catalogue number 06570 and BD/Pharmingen catalogue number 559565). DNA stains were Hoechst 33342 (Invitrogen, catalogue number H1399) propidium iodide (Invitrogen, catalogue number P1304MP) and ethidium bromide (Invitrogen, catalogue number 15585011). Fluorescent antibodies were from Invitrogen: goat anti-rabbit and goat anti-mouse Alexa 488, 555 and 647 (catalogue numbers A11001, A21422, A21235, A21238, A21428 and A21244).

Small molecule inhibitors. Brd4 bromodomain inhibitor (+)JQ1 and its inactive enantiomer (–)JQ1 were synthesized as described¹⁷ and were used at 250 nM. α -Amanitin (catalogue number A2263) and cycloheximide (catalogue number C4859) were from Sigma and were used at concentrations as indicated (α -amanitin 1–16 μ M; cycloheximide 35–560 μ M). UCN01 was from Sigma (catalogue number U6508) and was used at concentrations of 0.003–10 μ M. Caffeine was from Sigma (catalogue number C0750) and was used at concentrations of 10–25 mM. LBH589 was a gift from J. Bradner).

RNAi library. shRNA was applied to cells using a high-titre arrayed lenti-viral library maintained in the pLKO_TRC001 vector as described⁷.

Image-based screens. For shRNA screens and small molecule tests, human U2OS osteosarcoma cells (ATCC HTB-96) were grown in DMEM + Pen/Strep + 10% v/v FBS (complete media) at 37 °C in a 5% CO₂ atmosphere. All screens were performed at passage 10–15. Cells were tested for mycoplasma by PCR before seeding and infection. U2OS cells were seeded with a MicroFill (Biotek) in 384-well black, clear bottom plates (Greiner) at a density of 300 (shRNA) cells per well in 50 μ l of media, and allowed to attach overnight at 37 °C in a 5% CO₂ atmosphere. For shRNA screens, the media was exchanged the following day to complete media with 8 μ g ml^{–1} polybrene using a JANUS workstation (PerkinElmer). Virus infection was performed on an EP3 workstation (PerkinElmer) with 1.5 μ l of high-titre retrovirus. All plates had two wells infected with 1.5 μ l of control virus with shRNA directed against H2AX. Plates were centrifuged in a swinging-bucket rotor at 1180g for 30 min after infection and returned to the incubator overnight. The plates were then selected with 2.5 μ g ml^{–1} puromycin for 48 h, and allowed to proliferate in complete media for another 48 h, with media exchanges performed on the JANUS or RapidPlate (Qiagen) liquid handling workstations. Eight wells in each plate were not selected with puromycin. For small molecule testing, cells were plated at 500 cells per well in 384-well plates. The day after plating, small molecules at different concentrations in 100 nl DMSO were pin transferred to cells with a CyBio robot, and cells were propagated for 16 h. For both small molecule and shRNA screens, four plates were created in replicate for the time points outlined below. Four wells were left untreated in each plate, and received 25 mM caffeine in complete media 1 h before irradiation. All plates were treated with 10 Gy of 667 keV X-rays from a ¹³⁷Cs source in a Gammacell irradiator (Atomic Energy of Canada). A 0 h control plate was not irradiated. The plates were returned to the incubator and fixed with 4.4% w/v paraformaldehyde in phosphate-buffered saline (PBS) at 1, 6 and 24 h after irradiation. Plates were stored in PBS at 4 °C before staining. Fixed plates were washed three times with PBS and blocked with 24 μ l of GSDB (0.15% goat serum, 8.33% goat serum, 120 mM sodium phosphate, 225 mM NaCl) for 30 min. The 0, 1 and 6 h plates were incubated with 1:300 dilutions of GSDB of primary mouse monoclonal anti- γ H2AX (Ser 139), and rabbit polyclonal anti-pHH3 antibody. For the 24 h plates, we substituted 1:300 rabbit polyclonal anti-cleaved caspase 3 for the pHH3 antibody. All plates were incubated overnight at 4 °C, washed and stained with a secondary antibody mix containing 10 μ g ml^{–1} Hoescht 33342, 1:300 goat anti-mouse polyclonal-Alexa Fluor 488 and goat anti-rabbit polyclonal-Alexa Fluor 555 in GSDB. After a second overnight incubation at 4 °C, the plates were washed three times in PBS and stored in 50 μ l per well 50 μ M Trilox (Sigma) in PBS at 4 °C.

Imaging and image analysis. Plates were allowed to equilibrate to room temperature for 30 min and imaged on a Cellomics ArrayScan VTI automated microscope with a $\times 20$ objective lens. The acquisition parameters were the same for each

shRNA or chemical inhibitor. Six fields per well were imaged, with three channels/field (DAPI, fluorescein and rhodamine) for a total of 18 acquired images per well. Images were segmented and analysed with CellProfiler cell image analysis software. The imaging pipeline used to segment the images is available on request. Cell morphology and intensity data were acquired on a per image and per cell basis, and exported into a MySQL database. The data were visualized with SpotFire (TIBCO) and CellProfiler Analyst.

Immunofluorescence microscopy. U2OS cells were plated on number 1 glass coverslips (VWR) and were cultured in DMEM + Pen/Strep + 10% v/v FBS (complete media) at 37 °C in a 5% CO₂ atmosphere, then exposed to 10 Gy ionizing radiation from a ¹³⁷Cs source in a Gammacell irradiator (Atomic Energy of Canada), fixed in methanol and processed for immunofluorescence using the antibodies indicated above. Images were captured on a Zeiss Axiophot II microscope with a Hamamatsu CCD (charge-coupled device) camera and processed with OpenLab/Volocity software. We used CellProfiler (www.CellProfiler.org) or ImageJ software (<http://rsb.info.nih.gov/ij/>) for quantitative image analysis.

RT-PCR. Total RNA was extracted from 10⁶ U2OS cells expressing either control or Brd4-directed shRNA, with an RNeasy kit (Qiagen). Complementary DNA was generated with oligo(dT) primers with SuperScript reverse transcriptase (Invitrogen) according to the manufacturer's instructions. These complementary DNAs were used as templates for linear-range PCR amplification or quantitative real-time PCR with SYBR green master mix on an Applied Biosystems 7500 with the following primers: forward 5'-CTC CTC CTA AAA AGA CGA AGA-3' and reverse (pan-Brd4 isoform) 5'-TTC GGA GTC TTC GCT GTC AGA GGA G-3', (Brd4 isoform A) 5'-GCC CCT TCT TTT TTG ACT TCG GAG C-3', (Brd4 isoform B) 5'-GCC CTG GGG ACA CGA AGT CTC CAC T-3', (Brd4 isoform C) 5'-CCG TTT TAT TAA GAG TCC GTG TCC A-3', (CHEK2) forward 5'-ACAGATAAATAC CGAACATACAGC-3' and reverse 5'-GACGCGCTTTTCCTTCCCTACAA-3', and using (GAPDH) primers forward 5'-GATGCCCTGGAGGAAGTGCT-3' and reverse 5'-AGCAGGCACAA CACCAGTT-3' as control for normalization.

Expression profiling and analysis. Total RNA was collected from stable U2OS cells expressing Brd4 or control shRNA using RNeasy (Qiagen), labelled and analysed on the Affymetrix U133 Plus 2.0 array. Unsupervised clustering of expression data was performed using the R package pvclust. LIMMA²¹ was used to identify important changes in expression between Brd4 knockdown and control cells. Data were deposited in the US National Institutes of Health Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30700>).

Subcellular fractionation. U2OS cells expressing Flag-tagged Brd4 isoforms were lysed in hypotonic conditions (10 mM Hepes, 10 mM NaCl, 25 mM KCl, 1 mM MgCl₂, 0.1 mM EDTA, pH 7.4 with protease inhibitors) and subjected to flash freezing in liquid nitrogen 1 h after mock treatment or exposure to 10 Gy of ionizing radiation with a ¹³⁷Cs source in a Gammacell irradiator (Atomic Energy of Canada). Cells were thawed at room temperature and spun down at 10,000g for 10 min. The supernatant was saved as the cytoplasmic fraction and concentrated down using trichloroacetic acid precipitation and reconstituted in 2 \times Laemmli buffer. The pellet was re-suspended in high salt buffer (20 mM Hepes, 0.5 mM DTT, 1.5 mM MgCl₂, 0.1% Triton X-100, 1 M NaCl, pH 7.4 with protease inhibitors) and left on ice for 30 min followed by a high-speed spin at 100,000g for 30 min. The supernatant was saved as the high salt fraction and concentrated down using trichloroacetic acid precipitation and reconstituted in 2 \times Laemmli buffer. Sulphuric acid (0.4 N) was added to the high-speed pellet and left on ice for 30 min, followed by a high-speed spin at 14,000g for 10 min. The supernatant was saved as the acid fraction and concentrated down using trichloroacetic acid precipitation and reconstituted in 2 \times Laemmli buffer.

Western blotting and immunoprecipitation. Cells were treated with 10 Gy ionizing radiation with a ¹³⁷Cs source in a Gammacell irradiator (Atomic Energy of Canada). For whole cell lysates, cells were trypsinized and lysed in LB (4% SDS, 120 mM Tris, pH 6.8) with protease and phosphatase inhibitors (Complete mini EDTA-free and PhosSTOP, Roche Applied Science). For chromatin isolation, cells were trypsinized, re-suspended in low salt buffer (LSB: 10 mM Hepes 10 mM NaCl, 25 mM KCl, 1.0 mM MgCl₂, 0.1 mM EDTA, pH 7.4 + protease inhibitors, as above), flash-frozen in liquid N₂, thawed, pelleted at 10,000g for 10 min, re-suspended in high salt buffer (HSB: 20 mM Hepes, 1.0 M NaCl, 0.5 mM DTT, 1.5 mM MgCl₂, 0.1% Triton X-100 + protease inhibitors) for 45 min on ice, pelleted at 100,000g for 30 min., and proteins from the supernatant were precipitated with trichloroacetic acid. For immunoprecipitation, U2OS cells expressing Flag-tagged Brd4 isoforms were lysed in low salt buffer (50 mM Tris HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40 with protease inhibitors) and subjected to flash freezing in liquid nitrogen 1 h after mock treatment or irradiation. Cells were thawed at room temperature and spun down at 10,000g for 10 min. The supernatant was removed and saved as the pre-immunoprecipitation cytoplasmic fraction. The nuclear pellet was re-suspended in low salt buffer, tip sonicated at 4 °C (35% amplitude, pulse 5 s on and off for three cycles), and spun down at

14,000g for 10 min. The supernatant was collected as starting material for immunoprecipitation using M2 Flag beads (Sigma Aldrich) overnight at 4 °C. The beads were then spun down and the first supernatant saved as the unbound fraction. The beads were washed five times with low salt buffer and proteins were solubilized in 2× Laemmli buffer and boiled at 95 °C for 3 min before loading onto SDS–PAGE. Samples were processed after SDS–PAGE for gel band cutting and in gel tryptic digestion for mass spectrometry or western blotting to detect pulldown of the condensin II complex (SMC2 and SMC4 proteins) with Brd4 isoforms. SDS–PAGE and western blot was according to the methods of Laemmli and Towbin using either a Li-cor Odyssey scanner or horseradish-peroxidase-coupled secondary antibodies (Bio-Rad) and Western Lightning enhanced chemiluminescence (PerkinElmer) for visualization of bands.

Pulsed-field gel electrophoresis and micrococcal nuclease assay. For pulsed-field gel analysis, control and BRD4 knockdown cells were plated at 1×10^6 cells per plate, exposed to 10 Gy ionizing radiation with a ^{137}Cs source in a Gammacell irradiator (Atomic Energy of Canada) and collected at 0.5, 1, 2, 3 and 5 h. Cells were trypsinized, diluted to 2×10^6 cells and embedded in agarose plugs. The agarose plugs were exposed to Proteinase K (1 mg ml^{-1}) in 500 mM EDTA, 1% N-lauryl Sarcosyl, pH 8.0, for 48 h, washed $3 \times 1 \text{ h}$ with TE buffer, loaded onto a 0.675% agarose gel and separated under pulsed-field conditions with a Rotaphor 6.0 (Biometa). Nuclei from control and Brd4 knockdown cells were isolated by hypotonic lysis and micrococcal nuclease assays performed as described by Carey and Smale²².

Flow cytometry. U2OS cells were plated and transiently transfected GFP transgenes or siRNA as indicated, exposed to varying doses of ionizing radiation from a ^{137}Cs Gammacell irradiator source (Atomic Energy of Canada) and collected at varying times as indicated by fixation with 4% formaldehyde (cell-death measurements) or directly extracted with 100% ethanol (cell-cycle measurements), and processed for flow cytometry using the antibodies listed above. Data were analysed using FlowJo (www.flowjo.com) software.

Colony formation assays. Control and BRD4 knockdown cells were exposed to the indicated doses of ionizing radiation from a ^{137}Cs source in a Gammacell irradiator (Atomic Energy of Canada), or left untreated, trypsinized, counted and re-plated using serial dilutions. Colonies were propagated to the 10- to 15-cell stage (3–7 days), stained with Wright stain (Sigma) and counted with CellProfiler software or by averaging counts of ten fields from three independent observers using a dissection microscope to identify colonies of more than 15 cells.

Constructs, shRNA and siRNA, and transfection. Full-length constructs of Brd4 Isoform A (accession number NM_058243), B (accession number BC035266) and C (accession number NM_014299.2) were cloned into pEGFP-C1 (Clontech) and

pFLAG-CMV2 (Sigma) by PCR. Bromodomain mutations were introduced using quickchange (Stratagene) using PCR primers: 5'-AAA TTG TTA CAT CGC CAA CAA GCC TGG AGA TGA CGC AGT CTT AAT GGC AG-3' and 5'-CTG CCA TTA AGA CTG CGT CAT CTC CAG GCT TGT TGG CGA TGT AAC AAT TT-3'. Cells were transfected with Fugene 6 (Roche) according to the manufacturer's instructions. shRNA directed against Brd4 were from the TRC library (see Supplementary Table 1), or created in the mir30-based pMLP vector (a gift from M. Hemann) with primer 5'-TGC TGT TGA CAG TGA GCG AAG ACA CA-3' for Brd4. U2OS cell lines stably expressing this shRNA or control hairpins (ineffective hairpins directed against human sequences of BAD and PUMA) were created using puromycin selection at $2 \mu\text{g ml}^{-1}$. STEALTH siRNA against pan-isoform BRD4, SMC2 and control were purchased from Invitrogen. Custom Brd4 isoform-specific siRNA were synthesized from Dharmacon using the following sequences: isoform A specific 5'-GGG AGA AAG AGG AGC GUG AUU-3' and isoform B specific 5'-GCA CCA GUG GAG ACU UCG UUU-3'. siRNA against SMC2 was from Dharmacon. For siRNA experiments, cells were transfected with Lipofectamine RNAiMax (Invitrogen) according to the manufacturer's instructions.

Mass spectrometry. Proteins from the Brd4 co-immunoprecipitation were examined after SDS–PAGE by staining with Coomassie blue. Gel bands were excised, de-stained and processed for digestion with trypsin (Promega; $12.5 \text{ ng } \mu\text{l}^{-1}$ in 50 mM ammonium bicarbonate, pH 8.9). Peptides were loaded directly onto a column packed with C18 beads. The column was placed in-line with a tapered electrospray column packed with C18 beads on a Orbitrap XL mass spectrometer (Thermo Scientific). Peptides were eluted using a 120-min gradient (0–70% acetonitrile in 0.2 M acetic acid; 50 nl min^{-1}). Data were collected using the mass spectrometer in data-dependent acquisition mode to collect tandem mass spectra and examined using Mascot software (Matrix Science).

Network analysis. Protein–protein and kinase–substrate interactions relevant to DNA damage signalling were hand curated from primary literature available in PubMed using the initial keywords 'DNA damage', 'cell cycle checkpoint', 'chromatin structure', 'ATM/ATR', 'Chk1/Chk2' and 'SMC proteins', and following reference lists.

21. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
22. Carey, M. & Smale, S. T. Micrococcal nuclease–Southern blot assay: I. MNase and restriction digestions. *CSH Protoc.* **2007**, <http://dx.doi.org/10.1101/pdb.prot4890> (2007).

Chromosome-specific nonrandom sister chromatid segregation during stem-cell division

Swathi Yadlapalli^{1,2} & Yukiko M. Yamashita^{1,2,3}

Adult stem cells undergo asymmetric cell division to self-renew and give rise to differentiated cells that comprise mature tissue¹. Sister chromatids may be distinguished and segregated nonrandomly in asymmetrically dividing stem cells², although the underlying mechanism and the purpose it may serve remain elusive. Here we develop the CO-FISH (chromosome orientation fluorescence *in situ* hybridization) technique³ with single-chromosome resolution and show that sister chromatids of X and Y chromosomes, but not autosomes, are segregated nonrandomly during asymmetric divisions of *Drosophila* male germline stem cells. This provides the first direct evidence, to our knowledge, that two sister chromatids containing identical genetic information can be distinguished and segregated nonrandomly during asymmetric stem-cell divisions. We further show that the centrosome, SUN-KASH nuclear envelope proteins and *Dnmt2* (also known as *Mt2*) are required for nonrandom sister chromatid segregation. Our data indicate that the information on X and Y chromosomes that enables nonrandom segregation is primed during gametogenesis in the parents. Moreover, we show that sister chromatid segregation is randomized in germline stem cell overproliferation and dedifferentiated germline stem cells. We propose that nonrandom sister chromatid segregation may serve to transmit distinct information carried on two sister chromatids to the daughters of asymmetrically dividing stem cells.

The *Drosophila* male germline stem cell (GSC) system is an excellent model system for the study of asymmetric stem cell division. GSCs can be identified at single-cell resolution at the apical tip of the testis, where they attach to a cluster of somatic hub cells, a major component of the stem-cell niche⁴. GSCs divide asymmetrically by orienting the mitotic spindle perpendicular to the hub⁵. We showed previously that the mother centrosome is inherited by the GSCs⁶.

We adapted the CO-FISH (chromosome orientation fluorescence *in situ* hybridization) protocol, which allows strand-specific identification of sister chromatids³, combined with chromosome-specific probes⁷ (Fig. 1a). Using this method, we identified the sister chromatids of each chromosome in GSCs and their differentiating daughter gonialblasts (Fig. 1b and Supplementary Fig. 1). We found that sister chromatids of the Y chromosome are inherited with a strong bias during GSC division: In approximately 85% of cases, GSCs inherited the sister chromatid of the Y chromosome, whose template strand contains the (GTATT)₆ satellite (and thus hybridizes to the Cy3-(AATAC)₆ probe), and gonialblasts inherited the sister chromatid whose template contains the (AATAC)₆ sequence (and thus hybridizes to the Cy5-(GTATT)₆ probe; Fig. 1c, d). Using X-chromosome-specific probes, we found that the X chromosome shows a similar bias (Fig. 1e, f). Essentially the same results were obtained when the Cy5 probe for the X chromosome was replaced with a probe that is not complementary to the Cy3-labelled probe (Supplementary Fig. 2). Although both X and Y chromosomes show a similar bias in segregation (approximately 85:15), we found that the two chromosomes segregate independently of each other (Fig. 1g–i) (see Methods for details).

Two major scenarios can explain the observed bias of approximately 85:15. In the first scenario, approximately 85% of GSCs inherit the ‘red

strand’ (that is, the sister chromatid containing the template strand that hybridizes to Cy3 probes) with near 100% accuracy, whereas approximately 15% of GSCs inherit the ‘blue strand’ with near 100% accuracy. This would indicate that GSCs maintain particular strands of the X and Y chromosomes forever (‘immortal strands’). In the second scenario, each GSC inherits the ‘red strand’ with 85% probability and the ‘blue strand’ with 15% probability at each division. In this case, GSCs do not retain immortal strands; instead, the ‘template strands’ switch approximately once in every seven divisions (15% ≈ 1/6.7). To distinguish between these possibilities, we conducted a long-pulse experiment where flies were continuously exposed to 5-bromodeoxyuridine-containing medium (see Supplementary Fig. 3 for details). The results of this experiment clearly supported the second scenario.

In contrast to X and Y chromosomes, we found that the autosomes (chromosomes 2 and 3) do not show biased segregation (~50:50; Fig. 2). Consistent with previous reports that homologous chromosomes are paired, even in non-meiotic cells in *Drosophila*⁸, we observed that two autosome signals corresponding to homologous chromosomes were always juxtaposed to each other (Fig. 2a–d). In spite of the lack of biased segregation with regard to which strands are inherited by GSCs, cells always inherited two Cy3 signals or two Cy5 signals, the mechanism and significance of which remain elusive. It should be noted that the repeat sequences used as probes for chromosome 2 and 3 also exist on the Y chromosome⁹, yielding a third ‘lone’ signal in addition to the paired autosome signals. The identity of the lone signal was confirmed by combining autosome probes and a Y chromosome probe, 488-(AATAC)₆. The Y chromosome signal was always close to the lone signal (Fig. 2e, f). Importantly, the Y chromosome detected as a lone signal showed biased segregation, despite the fact that the paired autosome signals showed a random segregation pattern in the same set of samples (Fig. 2g). This result further confirms our observation that sister chromatids of the Y chromosome are segregated nonrandomly.

Although many studies have reported biased sister chromatid segregation, the genes responsible for biased segregation have never been described. We found that *centrosomin* (*cnn*), a core component of the pericentriolar material¹⁰, SUN domain protein KOI¹¹, and KASH domain protein KLAR¹² are required for biased sister chromatid segregation (Fig. 3, Supplementary Table 1). It is well established that the LINC (linker of nucleoskeleton and cytoskeleton) complex, composed of SUN- and KASH-domain proteins, tethers the nucleus to cytoskeletal components (such as microtubules, which in turn connect to the centrosome) via the nuclear envelope¹³. Thus, we speculate that specific sister chromatids are tethered to the mother centrosome of the GSC that is consistently located at the hub–GSC junction (see Fig. 4e).

We further found that sister chromatid segregation of X and Y chromosomes was randomized in *dnmt2* mutants (Supplementary Table 2a and Supplementary Fig. 4). Although some studies indicated that DNMT2 has DNA methyltransferase activity^{14,15}, other studies showed that it functions as an RNA methyltransferase¹⁶ and that DNA methylation is barely detectable in the *Drosophila* genome¹⁷. Therefore, the mechanism by which DNMT2 participates in nonrandom sister chromatid

¹Life Sciences Institute, Center for Stem Cell Biology, University of Michigan, Ann Arbor, Michigan 48109, USA. ²Department of Cell and Developmental Biology, School of Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA. ³Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, Michigan 48109, USA.

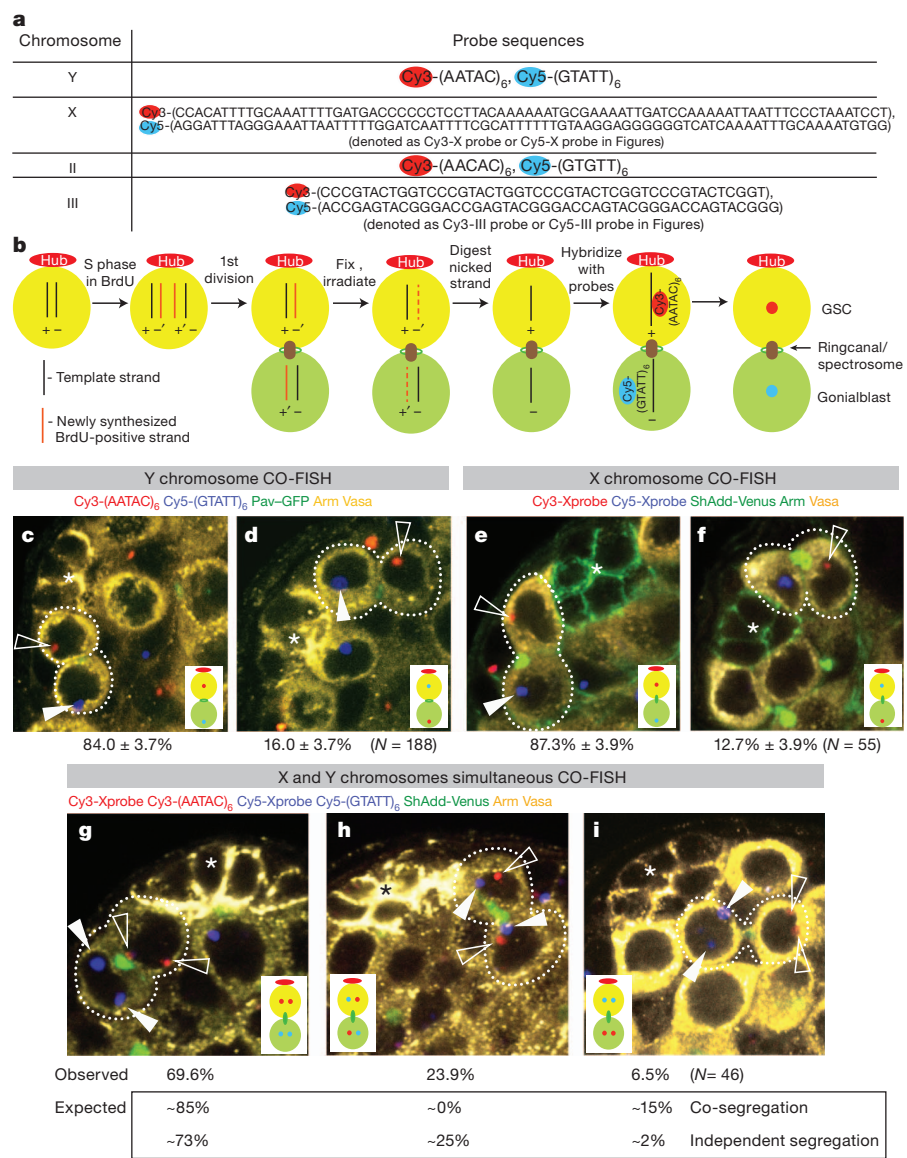


Figure 1 | Nonrandom segregation of Y and X chromosome strands during GSC divisions.
a, Chromosome-specific probes used in this study. **b**, Schematic diagram of the CO-FISH procedure. Cy3- and Cy5-labelled probes for the Y chromosome are shown as an example. Green fluorescent protein-labelled PAVAROTTI (PAV-GFP)²⁷ (midbody/ring canal), SH-ADD-Venus²⁸ or anti-ADD antibody (spectrosome) was used to identify GSC-gonialblast pairs. **c–i**, Representative images of CO-FISH results using Y chromosome probes (**c**, **d**), X chromosome probes (**e**, **f**), and both X and Y probes (**g–i**). Expected segregation patterns based on co-segregation versus random segregation are shown at the bottom of **g**, **h** and **i**. In all figures the Cy5 signal is indicated by solid arrowheads and the Cy3 signal by open arrowheads. An asterisk marks the position of the hub. N, number of GSC-gonialblast pairs scored. Data are presented as mean ± standard deviation.

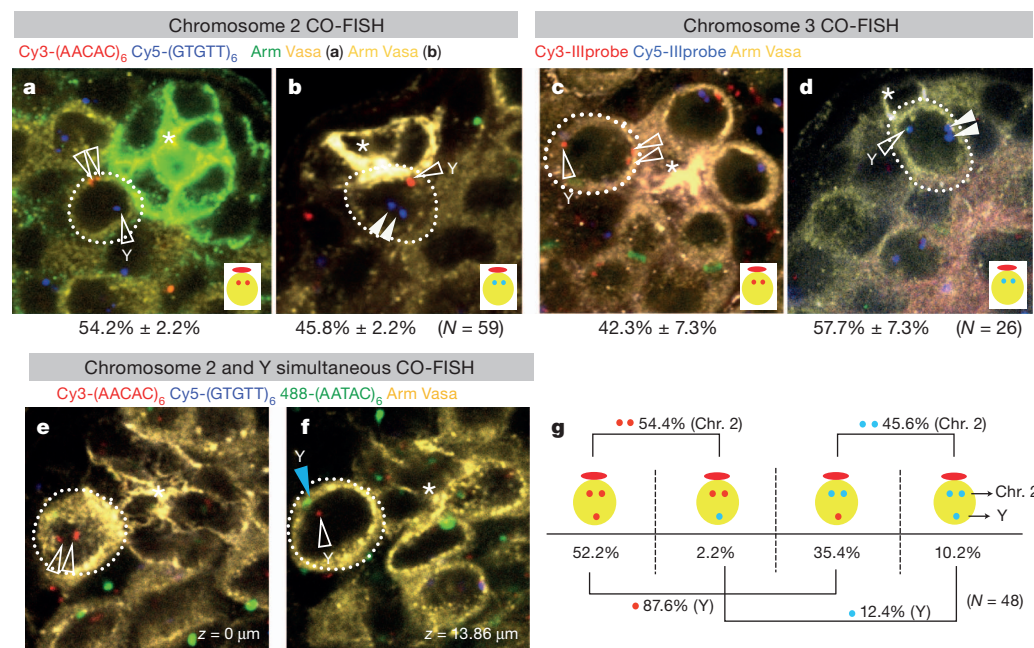


Figure 2 | Autosomes are randomly segregated during GSC divisions.
a–d, Representative images of CO-FISH results using chromosome 2 probes (**a**, **b**), and chromosome 3 probes (**c**, **d**). Lone signals that correspond to the Y chromosome are marked with 'Y'. N, number of GSCs scored. An asterisk marks the position of the hub. **e**, **f**, A representative image showing the lone signal of the (AACAC)₆ probe (open arrowheads) is close to the (AATAC)₆ signal (blue arrowhead). **g**, Summary of scoring results using chromosome 2 probes. Paired signals segregate randomly (Cy3-Cy3: Cy5-Cy5 = 54.4:45.6), whereas lone signals segregate nonrandomly (Cy3: Cy5 = 87.6:12.4). (AACAC)₆ and (AATAC)₆ sequences are on the same strand of the *Drosophila* Y chromosome.

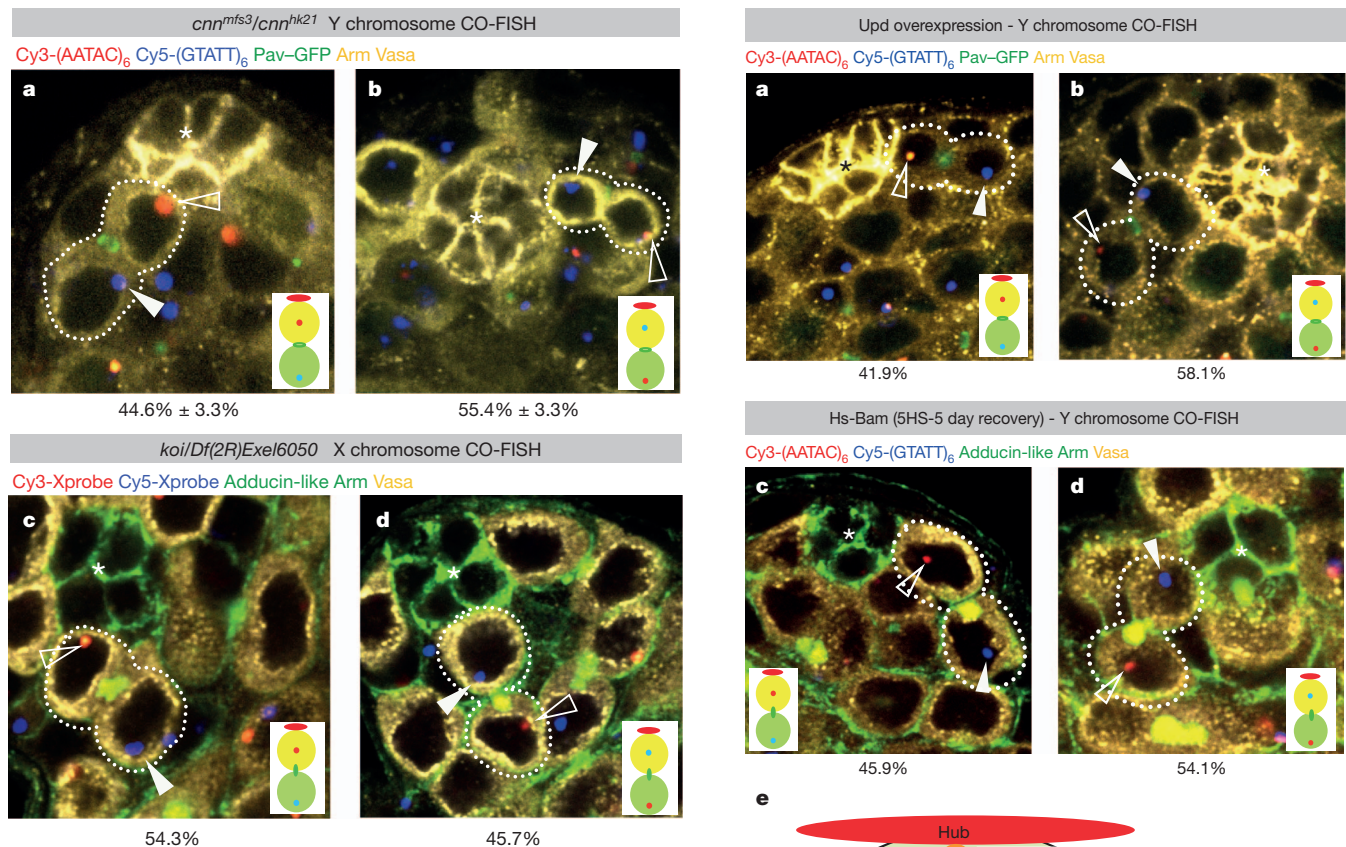


Figure 3 | *cnn*, *koi* and *klar* are required for nonrandom sister chromatid segregation. **a, b,** Representative images of Y chromosome CO-FISH in *cnn* mutant. Open arrowheads indicate the Cy3-(AATAC)₆ probe; closed arrowheads indicate the Cy5-(GTATT)₆ probe; asterisk indicates the hub. **c, d,** Representative images of X chromosome CO-FISH in *koi* mutant. Open arrowheads indicate the Cy3-X probe; closed arrowheads indicate the Cy5-X probe; asterisk indicates the hub.

segregation remains elusive. However, our analysis, using various crossing schemes (crosses of homozygous mother/father with heterozygous father/mother), indicates that DNMT2 confers heritable, DNA sequence-independent information on the X and Y chromosomes during gametogenesis in the parents, leading to nonrandom sister chromatid segregation of X and Y chromosomes in the GSCs of the progeny (Supplementary Table 2b). For example, in GSCs from flies that are genetically heterozygous (*dnmt2*^{+/-}), where the X chromosome is inherited from a mutant mother (*dnmt2*^{-/-}) and the Y chromosome from a heterozygous father (*dnmt2*^{+/-}), X chromosome segregation was randomized, whereas Y chromosome segregation remained nonrandom. These results suggest the striking possibility that the information that enables nonrandom sister chromatid segregation of X and Y chromosomes in adult stem cells is primed during gametogenesis in the parents, transmitted to the zygote on single X and Y chromosomes, and maintained through many cell divisions during embryogenesis and adult tissue homeostasis.

We found that sister chromatid segregation of X and Y chromosomes is randomized in GSC overproliferation induced by ectopic expression of UPD (also known as OS; Fig. 4a, b and Supplementary Table 3). UPD is a signalling ligand that is normally expressed exclusively in hub cells and activates the JAK-STAT pathway in GSCs and cyst stem cells to specify stem cell identity⁴. This finding indicates that nonrandom sister chromatid segregation is under the control of stem cell identity. However, it is unlikely that nonrandom sister chromatid segregation determines GSC identity, because the mutants defective in nonrandom segregation described above (*cnn*, *koi*, *klar*, *dnmt2*) do not show GSC overproliferation or depletion.

Figure 4 | Nonrandom segregation of Y and X chromosomes is disrupted in *upd*-overexpressing testes and dedifferentiated stem cells. **a, b,** Representative images of CO-FISH using the Y probe upon overexpression of UPD (nos-gal4 > UAS-UPD). For this experiment we limited our analysis to GSCs juxtaposed to hub cells, because GSCs located away from the hub do not have a spatial reference point for assessment of the sister chromatid segregation pattern. **N**, number of GSC-gonialblast pairs scored. An asterisk marks the position of the hub. **c, d,** Representative images of CO-FISH using the Y probe in dedifferentiated GSCs. Differentiation was induced by heat-shock treatment of hs-Bam flies followed by a 5-day recovery period²⁹. **e,** Model of nonrandom sister chromatid segregation (see text for details).

We also found that sister chromatid segregation is randomized in dedifferentiated GSCs (Fig. 4c, d and Supplementary Table 3). Partially differentiated germ cells can revert back to GSC identity to replenish the stem-cell pool^{18,19}. Although these dedifferentiated GSCs are apparently functional because they can produce differentiating spermatogonia and reconstitute spermatogenesis^{18,20}, they did not recover nonrandom sister chromatid segregation. This result may indicate that the information on X and Y chromosomes that allows nonrandom sister chromatid segregation is lost upon commitment to differentiation as a gonialblast. Consistent with our earlier observation that dedifferentiation increases during ageing²⁰, we found that nonrandom sister chromatid segregation was compromised during ageing (at day 30, 63:37 for the X chromosome (*N* = 35) and 68:32 for the Y chromosome (*N* = 28)).

This study provides the first evidence that adult stem cells can distinguish two sister chromatids, and further points to a model in which sister chromatids are distinctly recognized, leading to anchorage of particular strands to the mother centrosome through the SUN-KASH proteins

(Fig. 4e). Our data also indicate that nonrandom sister chromatid segregation does not necessarily mean that they are immortal²¹.

At present it is not clear why X and Y chromosomes segregate non-randomly. Considering the data presented in this study, we favour the possibility that certain epigenetic information is transmitted distinctively to GSCs and gonialblasts. Indeed, X and Y chromosomes are subject to various forms of epigenetic regulation, such as dosage compensation²² and male-specific meiotic sex chromosome inactivation²³. In addition, *Stellate*, a repetitive sequence that encodes a polypeptide known to reduce fertility, and Suppressor of *Stellate* (*Su(Ste)*), the Piwi-interacting RNA (piRNA) that suppresses *Stellate* expression, are located on the X and Y chromosomes, respectively^{24,25}. Intriguingly, we observed that *Stellate* is derepressed in mutants of *cnm*, *dnmt2*, *koi* and *klar* (Supplementary Fig. 5), although determination of whether derepression of *Stellate* is due to a failure in nonrandom sister chromatid segregation awaits future investigation. Not surprisingly, we found that the mutants in which *Stellate* is derepressed show reduced fertility (Supplementary Fig. 6).

Recently, it was shown that old versus new histones segregate asymmetrically during GSC divisions²⁶. Our study demonstrates that GSCs do not segregate old (immortal) DNA strands. Thus, the relationship between biased sister chromatid segregation and histone segregation remains elusive. In summary, our study presents the first evidence of chromosome-specific nonrandom sister chromatid segregation in adult stem cells and provides mechanistic insights into how cells segregate sister chromatids nonrandomly.

METHODS SUMMARY

For CO-FISH combined with immunofluorescence staining, newly eclosed flies (unless otherwise noted) were fed with 5-bromodeoxyuridine for ~10 h, followed by a period in non-5-bromodeoxyuridine medium (~10 h). The testes were then immunostained as described previously²⁰. Subsequently, testes were irradiated with ultraviolet light, followed by treatment with exonuclease III. Then, CO-FISH probes were hybridized to detect template strands.

Full Methods and any associated references are available in the online version of the paper.

Received 2 September 2012; accepted 20 March 2013.

Published online 5 May 2013.

- Morrison, S. J. & Kimble, J. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441**, 1068–1074 (2006).
- Tajbakhsh, S. & Gonzalez, C. Biased segregation of DNA and centrosomes: moving together or drifting apart? *Nature Rev. Mol. Cell Biol.* **10**, 804–810 (2009).
- Falconer, E. *et al.* Identification of sister chromatids by DNA template strand sequences. *Nature* **463**, 93–97 (2010).
- Fuller, M. T. & Spradling, A. C. Male and female *Drosophila* germline stem cells: two versions of immortality. *Science* **316**, 402–404 (2007).
- Yamashita, Y. M., Jones, D. L. & Fuller, M. T. Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome. *Science* **301**, 1547–1550 (2003).
- Yamashita, Y. M., Mahowald, A. P., Perlin, J. R. & Fuller, M. T. Asymmetric inheritance of mother versus daughter centrosome in stem cell division. *Science* **315**, 518–521 (2007).
- Dernburg, A. F. in *Drosophila Protocols* (eds Sullivan, W., Ashburner, M. & Hawley, R. S.) Ch. 2 (CSHL Press, 2000).
- Fung, J. C., Marshall, W. F., Dernburg, A., Agard, D. A. & Sedat, J. W. Homologous chromosome pairing in *Drosophila melanogaster* proceeds through multiple independent initiations. *J. Cell Biol.* **141**, 5–20 (1998).
- Makunin, I. V. *et al.* A novel simple satellite DNA is colocalized with the *Stalker* retrotransposon in *Drosophila melanogaster* heterochromatin. *Mol. Gen. Genet.* **261**, 381–387 (1999).

- Li, K. & Kaufman, T. C. The homeotic target gene *centrosomin* encodes an essential centrosomal component. *Cell* **85**, 585–596 (1996).
- Kracklauer, M. P., Banks, S. M., Xie, X., Wu, Y. & Fischer, J. A. *Drosophila klaroid* encodes a SUN domain protein required for Klarisicht localization to the nuclear envelope and nuclear migration in the eye. *Fly (Austin)* **1**, 75–85 (2007).
- Mosley-Bishop, K. L., Li, Q., Patterson, L. & Fischer, J. A. Molecular analysis of the *klarsicht* gene and its role in nuclear migration within differentiating cells of the *Drosophila* eye. *Curr. Biol.* **9**, 1211–1220 (1999).
- Razafsky, D. & Hodzic, D. Bringing KASH under the SUN: the many faces of nucleocytoplasmic connections. *J. Cell Biol.* **186**, 461–472 (2009).
- Phalke, S. *et al.* Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nature Genet.* **41**, 696–702 (2009).
- Kunert, N., Marhold, J., Stanke, J., Stach, D. & Lyko, F. A. Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development* **130**, 5083–5090 (2003).
- Schaefer, M. *et al.* RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* **24**, 1590–1595 (2010).
- Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919 (2010).
- Brawley, C. & Matunis, E. Regeneration of male germline stem cells by spermatogonial dedifferentiation *in vivo*. *Science* **304**, 1331–1334 (2004).
- Kai, T. & Spradling, A. Differentiating germ cells can revert into functional stem cells in *Drosophila melanogaster* ovaries. *Nature* **428**, 564–569 (2004).
- Cheng, J. *et al.* Centrosome misorientation reduces stem cell division during ageing. *Nature* **456**, 599–604 (2008).
- Yadlapalli, S., Cheng, J. & Yamashita, Y. M. *Drosophila* male germline stem cells do not asymmetrically segregate chromosome strands. *J. Cell Sci.* **124**, 933–939 (2011).
- Conrad, T. & Akhtar, A. Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nature Rev. Genet.* **13**, 123–134 (2012).
- Hense, W., Baines, J. F. & Parsch, J. X chromosome inactivation during *Drosophila* spermatogenesis. *PLoS Biol.* **5**, e273 (2007).
- Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**, 1017–1027 (2001).
- Tulin, A. V., Kogan, G. L., Filipp, D., Balakireva, M. D. & Gvozdev, V. A. Heterochromatic *Stellate* gene cluster in *Drosophila melanogaster*: structure and molecular evolution. *Genetics* **146**, 253–262 (1997).
- Tran, V., Lim, C., Xie, J. & Chen, X. Asymmetric division of *Drosophila* male germline stem cell shows asymmetric histone distribution. *Science* **338**, 679–682 (2012).
- Minestrini, G., Mathe, E. & Glover, D. M. Domains of the Pavarotti kinesin-like protein that direct its subcellular distribution: effects of mislocalisation on the tubulin and actin cytoskeleton during *Drosophila* oogenesis. *J. Cell Sci.* **115**, 725–736 (2002).
- Petrella, L. N., Smith-Leiker, T. & Cooley, L. The Ovhts polypeptide is cleaved to produce fusome and ring canal proteins required for *Drosophila* oogenesis. *Development* **134**, 703–712 (2007).
- Sheng, X. R., Brawley, C. M. & Matunis, E. L. Dedifferentiating spermatogonia outcompete somatic stem cells for niche occupancy in the *Drosophila* testis. *Cell Stem Cell* **5**, 191–203 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. Lyko, M. Schaefer, G. Reuter, P. Zamore, A. Aravin, D. Glover, L. Cooley, J. Kim, V. Gvozdev, M. Pia Bozzetti, the Bloomington *Drosophila* Stock Center and the Vienna *Drosophila* RNAi Center for reagents and helpful information, and Yamashita laboratory members for discussions. This study was supported by the University of Michigan (Life Sciences Institute and Office of the Provost and Executive Vice President for Academic Affairs) (to Y.M.Y.) and AHA (12PRE9630000) and NIH grants (1F31HD071727-01) (to S.Y.). Y.M.Y. is supported by the MacArthur Foundation.

Author Contributions S.Y. conceived the project and developed the single-chromosome CO-FISH protocol for *Drosophila* cells. S.Y. and Y.M.Y. designed and conducted experiments, interpreted the data, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.Y. (swathi@umich.edu) and Y.M.Y. (yukikomy@umich.edu).

METHODS

Fly husbandry. All fly stocks were raised on Bloomington Standard Media at 25 °C unless otherwise noted. The following fly stocks were used: Ubi-Pavarotti-GFP, SH-adducin-Venus, *cnm^{mfs3}/CyO*, *cnm^{HK21}/CyO*, *koi^{HIRKO80.w}*, Df(2R)Exel6050/CyO, *klar¹*, Df(3L)emc-E12, P(EP)M2^{G3429} (denoted *dnmt2^{G3429}* in the text), *dnmt2⁴⁹⁹*, *dnmt2¹⁴⁹*, Df(2L)ED775/CyO, hs-Bam, UAS-UPD/CyO, and nos-gal4. These stocks are described in FlyBase.

Combined immunofluorescence staining and CO-FISH. Newly eclosed adult flies (day 0) were fed food containing 5-bromodeoxyuridine (950 µl 100% apple juice, 7 µg agar, and 50 µl 100 mg ml⁻¹ 5-bromodeoxyuridine solution in a 1:1 mixture of acetone and DMSO) for approximately 10 h. After the feeding period, flies were transferred to regular fly food for approximately 10 h. Because the average GSC cell cycle length is 12 h, most GSCs undergo a single S phase followed by mitosis during our feeding procedure. GSCs that have undergone more or less than one S phase or mitosis were excluded from our analysis by limiting scoring to GSC–gonialblast pairs that have complementary CO-FISH signals in the GSC and gonialblast (that is, red signal in one cell, blue signal in the other). All possible scenarios are explained in Supplementary Fig. 1. Samples were dissected in 1× PBS, fixed for 30–60 min with 4% formaldehyde in PBS, permeabilized for at least 1 h in PBST (0.1% Triton X-100 in PBS) and incubated with primary antibodies overnight at 4 °C. Samples were then washed with PBST (20 min, three times), incubated overnight at 4 °C with Alexa Fluor-conjugated secondary antibodies (1:200; Molecular Probes), and washed again with PBST (20 min, three times). Samples were fixed for 10 min with 4% formaldehyde followed by three washes in PBST for 5 min each. Samples were then treated with RNase A (2 mg ml⁻¹ in water) for 10 min at 37 °C, washed with PBST for 5 min, and stained with 100 µl Hoechst 33258 (Sigma Aldrich) at 2 µg ml⁻¹ for 15 min at room temperature. The samples were then rinsed with 2× SSC, transferred to a tray, and irradiated with ultraviolet light in a UV Stratalinker 1800 (calculated dose, 5400 J m⁻²). Nicked 5-bromodeoxyuridine strands were digested with exonuclease III (New England Biolabs) at 3 U µl⁻¹ in buffer supplied by the manufacturer (50 mM Tris-HCl, 5 mM MgCl₂ and 5 mM dithiothreitol (DTT), pH 8.0) at 37 °C for 10 min. Samples were rinsed once with PBST for 5 min and then fixed in 4% formaldehyde in PBS for 2 min and washed three times for 5 min each in PBST. To allow gradual transition into 50% formamide/2× SSC, samples were incubated sequentially for a minimum of 10 min each in 20% formamide/2× SSC, 40% formamide/2×

SSC, and 50% formamide/2× SSC. The hybridization mixture consisted of 50% formamide, 2× SSC, 10% dextran sulphate, 0.5 µg ml⁻¹ Cy3-labelled probe, and 0.5 µg ml⁻¹ Cy5-labelled probe. Fluorescence-labelled probes were obtained from Integrated DNA Technologies. The hybridization solution was added to the samples and hybridization was carried out at 37 °C overnight. Using non-complementary pairs of probes for the X chromosome, we detected a similar bias in segregation pattern (Supplementary Fig. 2), excluding the possibility that annealing of complementary probes interferes with correct hybridization between the probes and the target sequences. Autosomal probes were denatured in hybridization solution at 65 °C for 3 min before hybridization. The samples were never heat-denatured. As a critical control, hub cells, which are predominantly quiescent and, thus, do not incorporate 5-bromodeoxyuridine, did not show any CO-FISH signal (evident in all images).

Following hybridization, samples were washed once in 50% formamide/2× SSC, once in 25% formamide/2× SSC, and three times in 2× SSC. Samples were then mounted in VECTASHIELD (H-1200, Vector Laboratories) and images were recorded using a Leica TCS SP5 confocal microscope with a 63× oil immersion objective (numerical aperture = 1.4) and processed using Adobe Photoshop software. The primary antibodies used were rabbit anti-Vasa (1:200; Santa Cruz Biotechnology), mouse anti-Adducin-like (1:20; developed by H. D. Lipshitz and obtained from the Developmental Studies Hybridoma Bank (DSHB)), mouse anti-Armadillo (1:20; developed by Eric Wieschaus and obtained from DSHB), rabbit anti-Stellate (1:1,000, a gift of P. Zamore³⁰). The secondary antibodies used were Alexa Fluor 594- and 488-conjugated secondary antibodies (1:200; Molecular Probes).

CO-FISH with both X and Y probes. The X and Y probes were labelled such that GSCs retain the Cy3 signal in ~85% of cases. If segregation of X and Y chromosomes is correlated, the probability that a GSC inherits two Cy3 signals will be approximately 85%, and that of inheriting two Cy5 signals will be approximately 15%, whereas there will be few instances where a GSC inherits one Cy3 and one Cy5 signal. In contrast, if the X and Y chromosomes segregate asymmetrically independently of each other, the probability of GSCs inheriting two Cy3 signals will be 72% (85% × 85%), that of inheriting two Cy5 signals will be 2% (15% × 15%), and that of inheriting one Cy3 and one Cy5 signal will be 26% (85% × 15% × 2).

30. Förstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).

TECHNOLOGY FEATURE

THE BIG CHALLENGES OF BIG DATA

As they grapple with increasingly large data sets, biologists and computer scientists uncork new bottlenecks.

EMBL-EBI



Extremely powerful computers are needed to help biologists to handle big-data traffic jams.

BY VIVIAN MARX

Biologists are joining the big-data club. With the advent of high-throughput genomics, life scientists are starting to grapple with massive data sets, encountering challenges with handling, processing and moving information that were once the domain of astronomers and high-energy physicists¹.

With every passing year, they turn more often to big data to probe everything from the regulation of genes and the evolution of genomes to why coastal algae bloom, what microbes dwell where in human body cavities

and how the genetic make-up of different cancers influences how cancer patients fare². The European Bioinformatics Institute (EBI) in Hinxton, UK, part of the European Molecular Biology Laboratory and one of the world's largest biology-data repositories, currently stores 20 petabytes (1 petabyte is 10^{15} bytes) of data and back-ups about genes, proteins and small molecules. Genomic data account for 2 petabytes of that, a number that more than doubles every year³ (see 'Data explosion').

This data pile is just one-tenth the size of the data store at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. Every

year, particle-collision events in CERN's Large Hadron Collider generate around 15 petabytes of data — the equivalent of about 4 million high-definition feature-length films. But the EBI and institutes like it face similar data-wrangling challenges to those at CERN, says Ewan Birney, associate director of the EBI. He and his colleagues now regularly meet with organizations such as CERN and the European Space Agency (ESA) in Paris to swap lessons about data storage, analysis and sharing.

All labs need to manipulate data to yield research answers. As prices drop for high-throughput instruments such as automated ►

► genome sequencers, small biology labs can become big-data generators. And even labs without such instruments can become big-data users by accessing terabytes (10^{12} bytes) of data from public repositories at the EBI or the US National Center for Biotechnology Information in Bethesda, Maryland. Each day last year, the EBI received about 9 million online requests to query its data, a 60% increase over 2011.

Biology data mining has challenges all of its own, says Birney. Biological data are much more heterogeneous than those in physics. They stem from a wide range of experiments that spit out many types of information, such as genetic sequences, interactions of proteins or findings in medical records. The complexity is daunting, says Lawrence Hunter, a computational biologist at the University of Colorado Denver. "Getting the most from the data requires interpreting them in light of all the relevant prior knowledge," he says.

That means scientists have to store large data sets, and analyse, compare and share them — not simple tasks. Even a single sequenced human genome is around 140 gigabytes in size. Comparing human genomes takes more than a personal computer and online file-sharing applications such as Dropbox.

In an ongoing study, Arend Sidow, a computational biologist at Stanford University in California, and his team are looking at specific changes in the genome sequences of tumours from people with breast cancer. They wanted to compare their data with the thousands of other published breast-cancer genomes and look for similar patterns in the scores of different cancer types. But that is a tall order: downloading the data is time-consuming, and researchers must be sure that their computational infrastructure and software tools are up to the task. "If I could, I would routinely look at all sequenced cancer genomes," says Sidow. "With the current infrastructure, that's impossible."

In 2009, Sidow co-founded a company called DNAnexus in Mountain View, California, to help with large-scale genetic analyses. Numerous other commercial and academic



Andreas Sundquist says amounts of data are now larger than the tools used to analyse them.

efforts also address the infrastructure needs of big-data biology. With the new types of data traffic jam honking for attention, "we now have non-trivial engineering problems", says Birney,

LIFE OF THE DATA-RICH

Storing and interpreting big data takes both real and virtual bricks and mortar. On the EBI campus, for example, construction is under way to house the technical command centre of ELIXIR, a project to help scientists across Europe safeguard and share their data, and to support existing resources such as databases and computing facilities in individual countries. Whereas CERN has one supercollider producing data in one location, biological research generating high volumes of data is distributed across many labs — highlighting the need to share resources.

Much of the construction in big-data biology is virtual, focused on cloud computing — in which data and software are situated in huge, off-site centres that users can access on demand, so that they do not need to buy their own hardware and maintain it on site. Labs that do have their own hardware can supplement it with the cloud and use both as needed. They can create virtual spaces for data, software and results that anyone can access, or they can lock the spaces up behind a firewall so that only a select group of collaborators can get to them.

Working with the CSC — IT Center for Science in Espoo, Finland, a government-run high-performance computing centre, the EBI is developing Embassy Cloud, a cloud-computing component for ELIXIR that offers secure data-analysis environments and is currently in its pilot phase. External organizations can, for example, run data-driven experiments in the EBI's computational environment, close to the data they need. They can also download data to compare with their own.

The idea is to broaden access to computing power, says Birney. A researcher in the Czech

Republic, for example, might have an idea about how to reprocess cancer data to help the hunt for cancer drugs. If he or she lacks the computational equipment to develop it, he or she might not even try. But access to a high-powered cloud allows "ideas to come from any place", says Birney.

Even at the EBI, many scientists access databases and software tools on the Web and through clouds. "People rarely work on straight hardware anymore," says Birney. One heavily used resource is the Ensembl Genome Browser, run jointly by the EBI and the Wellcome Trust Sanger Institute in Hinxton. Life scientists use it to search through, download and analyse genomes from armadillo to zebrafish. The main Ensembl site is based on hardware in the United Kingdom, but when users in the United States and Japan had difficulty accessing the data quickly, the EBI resolved the bottleneck by hosting mirror sites at three of the many remote data centres that are part of Amazon Web Services' Elastic Compute Cloud (EC2). Amazon's data centres are geographically closer to the users than the EBI base, giving researchers quicker access to the information they need.

More clouds are coming. Together with CERN and ESA, the EBI is building a cloud-based infrastructure called Helix Nebula — The Science Cloud. Also involved are infor-

"If I could, I would routinely look at all sequenced cancer genomes. With the current infrastructure, that's impossible."

mation-technology companies such as Atos in Bezons, France; CGI in Montreal, Canada; SixSq in Geneva; and T-Systems in Frankfurt, Germany.

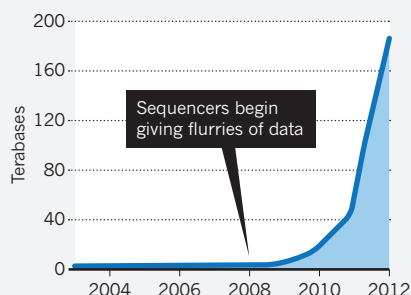
Cloud computing is particularly attractive in an era of reduced research funding, says Hunter, because cloud users do not need to finance or maintain hardware. In addition to academic cloud projects, scientists can choose from many commercial providers, such as Rackspace, headquartered in San Antonio, Texas, or VMware in Palo Alto, California, as well as larger companies including Amazon, headquartered in Seattle, Washington, IBM in Armonk, New York, or Microsoft in Redmond, Washington.

BIG-DATA PARKING

Clouds are a solution, but they also throw up fresh challenges. Ironically, their proliferation can cause a bottleneck if data end up parked on several clouds and thus still need to be moved to be shared. And using clouds means entrusting valuable data to a distant service provider who may be subject to power outages or other disruptions. "I use cloud services for many things, but always keep a local copy of scientifically important data and software," says Hunter. Scientists experiment with different constellations to

DATA EXPLOSION

The amount of genetic sequencing data stored at the European Bioinformatics Institute takes less than a year to double in size.



suit their needs and trust levels.

Most researchers tend to download remote data to local hardware for analysis. But this method is “backward”, says Andreas Sundquist, chief technology officer of DNAnexus. “The data are so much larger than the tools, it makes no sense to be doing that.” The alternative is to use the cloud for both data storage and computing. If the data are on a cloud, researchers can harness both the computing power and the tools that they need online, without the need to move data and software (see ‘Head in the clouds’). “There’s no reason to move data outside the cloud. You can do analysis right there,” says Sundquist. Everything required is available “to the clever people with the clever ideas”, regardless of their local computing resources, says Birney.

Various academic and commercial ventures are engineering ways to bring data and analysis tools together — and as they build, they have to address the continued data growth. Xing Xu, director of cloud computing at BGI (formerly the Beijing Genomics Institute) in Shenzhen, China, knows that challenge well. BGI is one of the largest producers of genomic data in the world, with 157 genome sequencing instruments working around the clock on samples from people, plants, animals and microbes. Each day, it generates 6 terabytes of genomic data. Every instrument can decode one human genome per week, an effort that used to take months or years and many staff.

DATA HIGHWAY

Once a genome sequencer has cranked out its snippets of genomic information, or ‘reads’, they must be assembled into a continuous stretch of DNA using computing and software. Xu and his team try to automate as much of this process as possible to enable scientists to get to analyses quickly.

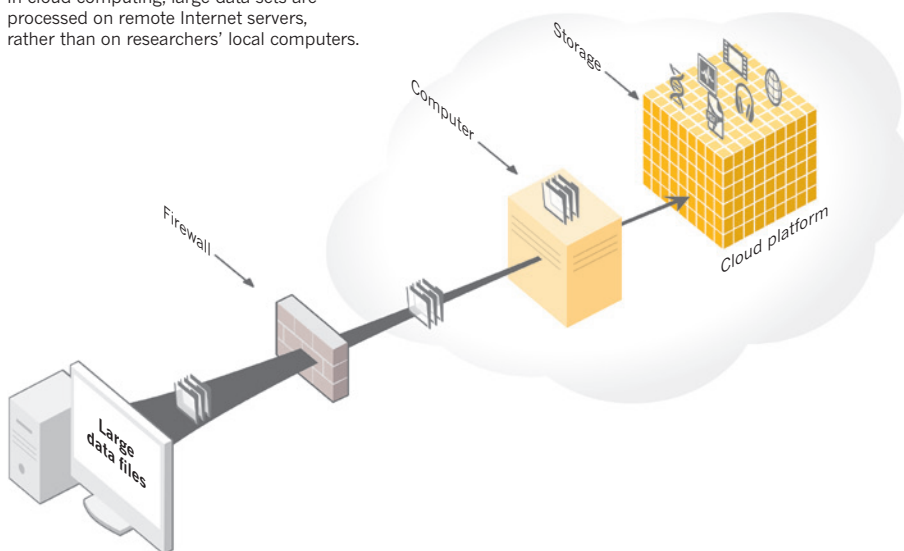
Next, either the reads or the analysis, or both, have to travel to scientists. Generally, researchers share biological data with their peers through public repositories, such as the EBI or ones run by the US National Center for Biotechnology Information in Bethesda, Maryland. Given the size of the data, this travel often means physically delivering hard drives — and risks data getting lost, stolen or damaged. Instead, BGI wants to use either its own clouds or others of the customer’s choosing for electronic delivery. But that presents a problem, because big-data travel often means big traffic jams.

Currently, BGI can transfer about 1 terabyte per day to its customers. “If you transfer one genome at a time, it’s OK,” says Xu. “If you sequence 50, it’s not so practical for us to transfer that through the Internet. That takes about 20 days.”

BGI is exploring a variety of technologies to accelerate electronic data transfer, among them *fasp*, software developed by Aspera in Emeryville, California, which helps to deliver

HEAD IN THE CLOUDS

In cloud computing, large data sets are processed on remote Internet servers, rather than on researchers’ local computers.



data for film-production studios and the oil and gas industry as well as the life sciences. In an experiment last year, BGI tested a *fasp*-enabled data transfer between China and the University of California, San Diego (UCSD). It took 30 seconds to move a 24-gigabyte file. “That’s really fast,” says Xu.

Data transfer with *fasp* is hundreds of times quicker than methods using the normal Internet protocol, says software engineer Michelle Munson, chief executive and co-founder of Aspera. However, all transfer protocols share challenges associated with transferring large, unstructured data sets.

The test transfer between BGI and UCSD was encouraging because Internet connections between China and the United States are “riddled with challenges” such as variations in signal strength that interrupt data transfer, says Munson. The protocol has to handle such road bumps and ensure speedy transfer, data integrity and privacy. Data transfer often slows

“There’s no reason to move data outside the cloud. You can do analysis right there.”

when the passage is bumpy, but with *fasp* it does not. Transfers can fail when a file is partially sent; with ordinary Internet connections, this relaunches the entire transfer. By contrast, *fasp* restarts where the previous transfer stopped. Data that are already on their way do not get resent, but continue on their travels.

Xu says that he liked the experiment with *fasp*, but the software does not solve the data-transfer problem. “The main problem is not technical, it is economical,” he says. BGI would need to maintain a large Internet connection bandwidth for data transfer, which would be prohibitively expensive, especially given that Xu and his team do not send out big data in a continuous flow. “If we only transfer

periodically, it doesn’t make any economic sense for us to have this infrastructure, especially if the user wants that for free,” he says.

Data-sharing among many collaborators also remains a challenge. When BGI uses *fasp* to share data with customers or collaborators, it must have a software licence, which allows customers to download or upload the data for free. But customers who want to share data with each other using this transfer protocol will need their own software licences. Putting the data on the cloud and not moving them would bypass this problem; teams would go to the large data sets, rather than the other way around. Xu and his team are exploring this approach, alongside the use of Globus Online, a free Web-based file-transfer service from the Computation Institute at the University of Chicago and the Argonne National Laboratory in Illinois. In April, the Computation Institute team launched a genome-sequencing-analysis service called Globus Genomics on the Amazon cloud.

Munson says that Aspera has set up a pay-as-you-go system on the Amazon cloud to address the issue of data-sharing. Later this year, the company will begin selling an updated version of its software that can be embedded on the desktop of any kind of computer and will let users browse large data sets much like a file-sharing application. Files can be dragged and dropped from one location to another, even if those locations are commercial or academic clouds.

The cost of producing, acquiring and disseminating data is decreasing, says James Taylor, a computational biologist at Emory University in Atlanta, Georgia, who thinks that “everyone should have access to the skills and tools” needed to make sense of all the information. Taylor is a co-founder of an academic platform called Galaxy, which lets scientists analyse their data and share software tools and workflows for free. Through

Web-based access to computing facilities at Pennsylvania State University (PSU) in University Park, scientists can download Galaxy's platform of tools to their local hardware, or use it on the Galaxy cloud. They can then plug in their own data, perform analyses and save the steps in them, or try out workflows set up by their colleagues.

Spearheaded by Taylor and Anton Nekrutenko, a molecular biologist at PSU, the Galaxy project draws on a community of around 100 software developers. One feature is Tool Shed, a virtual area with more than 2,700 software tools that users can upload, try out and rate. Xu says that he likes the collection and its ratings, because without them, scientists must always check if a software tool actually runs before they can use it.

KNOWLEDGE IS POWER

Galaxy is a good fit for scientists with some computing know-how, says Alla Lapidus, a computational biologist in the algorithmic biology lab at St Petersburg Academic University of the Russian Academy of Sciences, which is led by Pavel Pevzner, a computer scientist at UCSD. But, she says, the platform might not be the best choice for less tech-savvy researchers. When Lapidus wanted to disseminate the software tools that she developed, she chose to put them on DNAnexus's newly launched second-generation commercial cloud-based analysis platform.

That platform is also designed to cater to non-specialist users, says Sundquist. It is possible for a computer scientist to build his or her own biological data-analysis suite with software tools on the Amazon cloud, but DNAnexus uses its own engineering to help researchers without the necessary computer skills to get to the analysis steps.

Catering for non-specialists is important when developing tools, as well as platforms. The Biomedical Information Science and Technology Initiative (BISTI) run by the US National Institutes of Health (NIH) in Bethesda, Maryland, supports development of new computational tools and the maintenance of existing ones. "We want a deployable tool," says Vivien Bonazzi, programme director in computational biology and bioinformatics at the National Human Genome Research Institute, who is involved with BISTI. Scientists who are not heavy-duty informatics types need to be able to set up these tools and use them successfully, she says. And it must be possible to scale up tools and update them as data volume grows.

Bonazzi says that although many life scientists have significant computational skills, others do not understand computer lingo enough to know that in the tech world, Python is not a snake and Perl is not a gem (they are programming languages). But even if biologists can't develop or adapt the software, says Bonazzi, they have a place in big-data science. Apart from anything else, they can offer

valuable feedback to their computationally fluent colleagues because of different needs and approaches to the science, she says.

Increasingly, big genomic data sets are being used in biotechnology companies, drug firms and medical centres, which also have specific needs. Robert Mulroy, president of Merrimack Pharmaceuticals in Cambridge, Massachusetts, says that his teams handle mountains of data that hide drug candidates. "Our view is that biology functions through systems dynamics," he says.

Merrimack researchers focus on interrogating molecular signalling networks in the healthy body and in tumours, hoping to find new ways to corner cancer cells. They generate and use large amounts of information from the genome and other factors that drive a cell to become cancerous, says Mulroy. The company stores its data and conducts analysis on its own computing infrastructure, rather than a cloud, to keep the data private and protected.

Drug developers have been hesitant about cloud computing. But, says Sundquist, that fear is subsiding in some quarters: some companies that have previously avoided clouds because of security problems are now exploring them. To assuage these users' concerns, Sundquist has engineered the DNAnexus cloud to be compliant with US and European regulatory guidelines. Its security features include encryption for biomedical information, and logs to allow users to address potential queries from auditors such as regulatory agencies, all of which is important in drug development.

CHALLENGES AND OPPORTUNITIES

Harnessing powerful computers and numerous tools for data analysis is crucial in drug discovery and other areas of big-data biology. But that is only part of the problem. Data and tools need to be more than close — they must talk to one another. Lapidus says that results produced by one tool are not always in a format that can



Arend Sidow wants to move data mountains without feeling pinched by infrastructure.

be used by the next tool in a workflow. And if software tools are not easily installed, computer specialists will have to intervene on behalf of those biologists without computer skills.

Even computationally savvy researchers can get tangled up when wrestling with software and big data. "Many of us are getting so busy analysing huge data sets that we don't have time to do much else," says Steven Salzberg, a computational biologist at Johns Hopkins University in Baltimore, Maryland. "We have to spend some of our time figuring out ways to make the analysis faster, rather than just using the tools we have."

Yet other big-data pressures come from the need to engineer tools for stability and longevity. Too many software tools crash too often. "Everyone in the field runs into similar problems," says Hunter. In addition, research teams may not be able to acquire the resources they need, he says, especially in countries such as the United States, where an academic does not gain as much recognition for software engineering as for publishing a paper. With its dedicated focus on data and software infrastructure designed to serve scientists, the EBI offers an "interesting contrast to the US model," says Hunter.

US funding agencies are not entirely ignoring software engineering, however. In addition to BISTI, the NIH is developing Big Data to Knowledge (BD2K), an initiative focused on managing large data sets in biomedicine, with elements such as data handling and standards, informatics training and software sharing. And as the cloud emerges as a popular place to do research, the agency is also reviewing data-use policies. An approved study usually lays out specific data uses, which may not include placing genomic data on a cloud, says Bonazzi. When a person consents to have his or her data used in one way, researchers cannot suddenly change that use, she says. In a big-data age that uses the cloud in addition to local hardware, new technologies in encryption and secure



Various data-transfer protocols handle problems in different ways, says Michelle Munson.

transmission will need to address such privacy concerns.

Big data takes large numbers of people. BGI employs more than 600 engineers and software developers to manage its information-technology infrastructure, handle data and develop software tools and workflows. Scores of informaticians look for biologically relevant messages in the data, usually tailored to requests from researchers and commercial customers, says Xu. And apart from its stream of research collaborations, BGI offers a sequencing and analysis service to customers. Early last year, the institute expanded its offerings with a cloud-based genome-analysis platform called EasyGenomics.

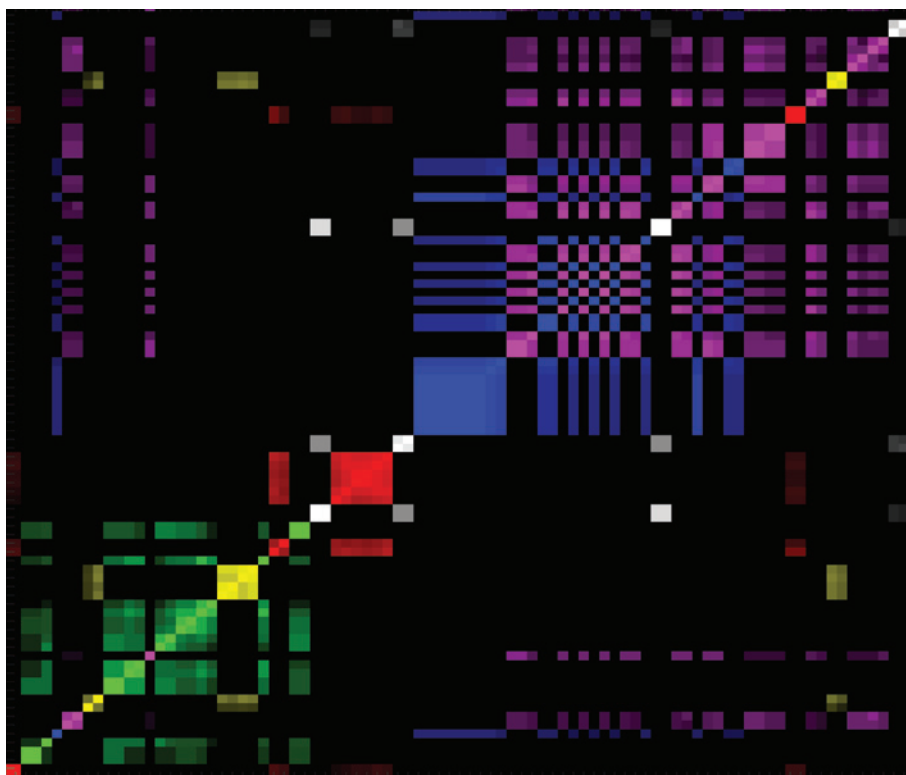
In late 2012, it also bought the faltering US company Complete Genomics (CG), which offered human genome sequencing and analysis for customers in academia or drug discovery. Although the sale dashed hopes for earnings among CG's investors, it doesn't seem to have dimmed their view of the prospects for sequencing and analysis services. "It is now just a matter of time before sequencing data are used with regularity in clinical practice," says one investor, who did not wish to be identified. But the sale shows how difficult it can be to transition ideas into a competitive marketplace, the investor says.

When tackling data mountains, BGI uses not only its own data-analysis tools, but also some developed in the academic community. To ramp up analysis speed and capacity as data sets grow, BGI assembled a cloud-based series of analysis steps into a workflow called Gaea, which uses the Hadoop open-source software framework. Hadoop was written by volunteer developers from companies and universities, and can be deployed on various types of computing infrastructure. BGI programmers built on this framework to instruct software tools to perform large-scale data analysis across many computers at the same time.

If 50 genomes are to be analysed and the results compared, hundreds of computational steps are involved. The steps can run either sequentially or in parallel; with Gaea, they run in parallel across hundreds of cloud-based computers, reducing analysis time rather like many people working on a single large puzzle at once. The data are on the BGI cloud, as are the tools. "If you perform analysis in a non-parallel way, you will maybe need two weeks to fully process those data," says Xu. Gaea takes around 15 hours for the same number of data.

To leverage Hadoop's muscle, Xu and his team needed to rewrite software tools. But the investment is worth it because the Hadoop framework allows analysis to continue as the

"The cultural baggage of biology that privileges data generation over all other forms of science is holding us back."



A simplified array of breast-cancer subtypes, produced by researchers at Merrimack Pharmaceuticals, who use their own computational infrastructure to hunt for new cancer drugs.

data mountains grow, he says.

They are still ironing out some issues with Gaea, comparing its performance on the cloud with its performance on local infrastructure. Once testing is complete, BGI plans to mount Gaea on a cloud such as Amazon for use by the wider scientific community.

Other groups are also trying to speed up analysis to cater to scientists who want to use big data. For example, Bina Technologies in Redwood City, California, a spin-out from Stanford University and the University of California, Berkeley, has developed high-performance computing components for its genome-analysis services. Customers can buy the hardware, called the Bina Box, with software, or use Bina's analysis platform on the cloud.

FROM VIVO TO SILICO

Data mountains and analysis are altering the way science progresses, and breeding biologists who get neither their feet nor their hands wet. "I am one of a small original group who made the first leap from the wet world to the *in silico* world to do biology," says Marcie McClure, a computational biologist at Montana State University in Bozeman. "I never looked back,"

During her graduate training, McClure analysed a class of viruses known as retroviruses in fish, doing the work of a "wet-worlder", as she calls it. Since then, she and her team have discovered 11 fish retroviruses without touching water in lake or lab, by analysing genomes computationally and in ways that others had not. She has also developed software tools to find such

viruses in the genomes of other species, including humans. Her work generates terabytes of data, which she shares with other researchers.

Given that big-data analysis in biology is incredibly difficult, Hunter says, open science is becoming increasingly important. As he explains, researchers need to make their data available to the scientific community in a useful form, for others to mine. New science can emerge from the analysis of existing data sets: McClure generates some of her findings from other people's data. But not everyone recognizes that kind of biology as an equal. "The cultural baggage of biology that privileges data generation over all other forms of science is holding us back," says Hunter.

A number of McClure's graduate students are microbial ecologists, and she teaches them how to rethink their findings in the face of so many new data. "Before taking my class, none of these students would have imagined that they could produce new, meaningful knowledge, and new hypotheses, from existing data, not their own," she says. Big data in biology add to the possibilities for scientists, she says, because data sit "under-analysed in databases all over the world". ■

Vivien Marx is technology editor at *Nature* and *Nature Methods*.

1. Mattmann, C. *Nature* **493**, 473–475 (2013).
2. Greene, C. S. & Troyanskaya, O. G. *PLoS Comput. Biol.* **8**, e1002816 (2012).
3. EMBL–European Bioinformatics Institute EMBL-EBI Annual Scientific Report 2012 (EMBL-EBI, 2013).

CAREERS

EMPLOYMENT Graduate-degree holders have higher pay and more jobs **p.263**

MEDICINE Survey finds high levels of burnout among oncologists **p.263**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



DEBRA HUGHES/SHUTTERSTOCK

ECOLOGY

Conservation in captivity

Zoos provide an opportunity to work on crucial issues of biodiversity while reaching out to the public.

BY AMANDA MASCARELLI

Barbara Durrant heard about San Diego Zoo's reproductive-research department while she was pursuing her doctorate in reproductive physiology in the late 1970s. "I wrote to the founder and got a wonderful letter back saying, 'Yes, we're starting this new

research effort here. When you finish your PhD, get back in touch with me," recalls Durrant. In 1979, she began a two-year postdoc at the zoo in California.

Looking for a second project towards the end of her stint, Durrant began collecting viable eggs, sperm and embryos from animals that had died, and storing them in the

facility's Frozen Zoo, one of the world's first major collections of cryopreserved cells from zoo animals. In 1980, she initiated the Germ-plasm Repository — a collection of frozen reproductive cells from endangered species that capture genetic diversity, allowing it to be reintroduced into gene pools. In so doing, she helped to launch the field of gamete research. After her postdoc ended later that year, the zoo offered Durrant a permanent research position. Now director of reproductive physiology at San Diego Zoo Global, the conservation organization that runs the zoo, Durrant heads a team that designs reproductive-research programmes for rare and endangered species including giant pandas, rhinoceroses and Przewalski's horses. "The greater scientific community is coming to understand the importance of genetic diversity," says Durrant. "And zoos harbour the greatest genetic diversity anywhere outside of the natural world."

In the past few decades, zoos and aquariums around the globe have transformed themselves. No longer just family destinations and collections of rare, threatened and endangered animals, they are also research institutions with conservation and science at the core of their mission. Zoos are well positioned to manage populations of animals whose numbers are rapidly dwindling in their natural habitat, and, in some cases, to reintroduce them into the wild. And although they have tended to emphasize captive-breeding programmes, zoos are becoming increasingly focused on field-based research and on saving species in the wild.

CALL OF THE WILD

Research positions involving conservation at zoos and aquariums are still relatively sparse. But many scientists find such jobs deeply satisfying. The research is mission-driven and aimed at solving immediate problems, so zoological facilities tend to attract scientists who embrace an applied approach, says Allison Alberts, chief conservation and research officer at San Diego Zoo Global.

"I always thought I was going to end up in the traditional academic environment," says Alberts. "I value academic research very much. But I wanted to do something more immediate. I saw a crisis in the world that needed to be addressed now. I felt like, 'I don't have the luxury to wait and see if my research is going to be relevant 30 years from now — I want to be doing something that's solving the conservation problem today.' And the zoo gave me the opportunity to do that." ►

► Like Durrant, Alberts joined the San Diego Zoo as a postdoc, and ended up forging her career there. Whereas some positions at zoos and aquariums involve only research, others may require engaging with the public and overseeing staff and projects. In addition to coordinating all of San Diego Zoo Global's research initiatives in areas such as sustainable populations, restoration biology and habitat conservation, Alberts helps to raise the funding that supports the zoo's conservation efforts. She misses hands-on research, but says that being part of the zoo's conservation initiatives provides a "whole different type of satisfaction".

With one of the largest zoological conservation programmes in the world, the San Diego Zoo employs more than 200 researchers, including 12 field-conservation postdocs. It has research projects in 38 countries and an annual conservation budget of US\$15 million, of which \$6 million comes from grants and government contracts, and the rest from donations and zoo operations.

Zoos that employ such large numbers of researchers are rare. However, many have robust conservation-science programmes; they include the Calgary Zoo in Canada, the Smithsonian Institution's National Zoo in Washington DC, Antwerp Zoo in Belgium and London Zoo. In addition to postdoc positions, researchers may find work as technicians, field and lab managers, educators or scientists leading their own research programmes at the zoo or in the field. Many, including Durrant and Alberts, are adjunct or full professors at nearby universities, enabling them to mentor students directly and to forge collaborations with academic researchers. And scientists with PhDs are sometimes employed as curators in a specific area such as reptiles or birds.

ALL CREATURES GREAT AND SMALL

Although in the past zoos have not tended to be seen as research centres, that is changing. "Within more traditional academia, I think it's quite easy to dismiss zoos and aquariums as a place where you could do real science," says Jackie Ogden, vice-president of animals, science and environment at Walt Disney Parks and Resorts, who is based at Disney's Animal Kingdom in Orlando, Florida. Ogden says that Disney researchers have been involved in more than 300 scientific articles in the past 15 years. Her team includes 14 PhD students, most of them active in conservation research, she says. In one project, researchers monitor sea-turtle nesting on the central Florida coast in collaboration with local universities and state wildlife agencies. Disney researchers have contributed to rehabilitation of more than 350 sea turtles over the past 20 years, says Ogden.

Aquariums have also grown into strong conservation-research centres. The Tennessee Aquarium Conservation Institute, the research arm of the Tennessee Aquarium in Chattanooga, is involved in restoration and

reintroduction of two imperilled fish species — lake sturgeon and southern Appalachian brook trout — to the Tennessee River system. Anna George, the institute's director and chief research scientist, says that the job gives her the opportunity to put conservation principles into practice. With a PhD in conservation genetics of freshwater species, she has a deep understanding of field-based genetic diversity. Her work lets her apply that knowledge while collaborating with others who have expertise in raising fish in captivity. "We can make sure that we're really recovering a species with the ability to adapt, not just putting individuals into a river," she says.

WALK WITH THE ANIMALS

As zoos and aquariums become more conservation-oriented, their research increasingly focuses on animals in their natural habitats. As a result, opportunities are growing for researchers to work with plants and animals in the field, says Ron Swaisgood, director of applied animal ecology at San Diego Zoo Institute for Conservation Research. "Zoos are in the process of reinventing themselves," he says. "People don't think of plant ecology as being a zoo research programme — but it is."

Swaisgood and others think that such jobs, including field research, will continue



"I think it's quite easy to dismiss zoos and aquariums as a place where you could do real science."

Jackie Ogden

to grow as zoos become focused on conservation, pooling resources from donations and external grants from local, state and federal regulatory agencies. In 2011, facilities accredited by the US Association of Zoos and Aquariums in Silver Spring, Maryland, spent US\$160 million on 2,670 research and conservation projects in more than 100 countries, up from US\$134 million in 2010, according to the association's most

recent annual report on conservation science.

In 2009, the European Association of Zoos and Aquaria (EAZA), based in Amsterdam, estimated that it provides €30 million (US\$39 million) per year in paid staff time and costs for zoological research. It also reported that 1,400–1,500 people conduct or facilitate research as part of their jobs in zoos and aquariums in Europe. This July, the EAZA will launch its own online, open-access publication, *The Journal of Zoo and Aquarium Research*, to provide more outlets for zoo-oriented science.

Research in zoos can be quite different from field research, says Lesley Dickie, executive director of the EAZA. For instance, she says,

DISNEY



A researcher from San Diego Zoo in California tracks a koala on St Bees Island, Australia.

zoo research might focus on animal behaviours that are not seen in the wild because they are very hard to observe. But if the research concentrates on a highly threatened species, sample sizes in both the wild and captivity might be very small, making zoo work that much more relevant to 'real-world' circumstances, and more valuable. "As the wild gets more and more pressurized, I think that some of the things we're learning about small-population management in zoos will be more and more applied to the wild," says Dickie.

HUMANS AND OTHER ANIMALS

Myriad skill sets can open doors to work in zoos and aquariums. Basic research in areas such as animal behaviour or reproductive biology continues to be important, says Durrant, and training in genetics, wildlife disease and conservation education is also valued. It is not necessary to have worked with exotic animals or in zoos previously, she notes: basic-research training with model species in universities is sufficient. "Get the strongest solid foundation you can get and that you can apply to conservation."

However, some experience at a zoo or aquarium, even as a volunteer, can make the transition easier. While doing her PhD at Saint Louis University in Missouri, George began working in the education department at Saint Louis Zoo, leading overnight and summer-camp education programmes. That experience was key to her being hired at the Tennessee Aquarium. "They knew I already understood the culture and goals of zoos and aquariums and the informal science-education part of that," she says. "So even if it's volunteering or serving as a keeper, that first step into it makes it a lot easier to get a job later."

Scientists interested in zoo work would do well to supplement their training with other skills related to conservation. Classes in non-profit management and fund-raising

can help. And George advises that researchers get comfortable with outreach, including the art of educating donors about their research.

"We need people who are limber enough to move between field and zoo," says John Fraser, a conservation psychologist who is president of the New Knowledge Organization, a social-science think tank based in New York. "It's the ability to have a foot in both worlds, with the authority of the field biologist and the access of the zoo biologist." He suggests pairing a field-biology degree with a minor in community organizing, organizational psychology or advocacy.

Regardless of the academic path, the ability to work with people — not just animals — is crucial. "The outreach I do ranges from elementary-school students to politicians to journalists and everything in between," says George. "Each programme is different. You have to be comfortable being flexible." Rachel Lowry, director of wildlife conservation and science at Zoos Victoria in Melbourne, Australia, finds that her most profound experiences come from engaging with audiences and helping to influence people's behaviour. "Zoos are really powerful conservation organizations because they have an enormous reach, and because they are entrusted with these incredible animals within their care," says Lowry. "To have an orang-utan stand behind you while you give a talk, and you say, 'Who here pledges to purchase only certified sustainable palm oil?' and an orang-utan raises its hand — it's very moving. Everyone standing in front of that orang-utan who has come to connect with it emotionally suddenly raises their hand and says, 'Yeah, I don't want that species to go extinct because of the food that I choose.' It's a really powerful role." ■

Amanda Mascarelli is a freelance writer based in Denver, Colorado.

EMPLOYMENT

On the job

US graduate-degree holders aged 30–54 with a background in life or physical sciences had an unemployment rate of 2.1% and a median salary of US\$90,000 in 2010–11, according to an analysis of census data. *Hard Times 2013: College Majors, Unemployment and Earnings*, released on 29 May by Georgetown University Center on Education and the Workforce in Washington DC, found that life- or physical-science graduates in the same age range with only a bachelor's degree had 4.8% unemployment and a median salary of \$60,000. With research jobs scarce, many science-graduate-degree holders work in secondary education, or in non-research posts in industries such as pharmaceuticals or aerospace, notes co-author Anthony Carnevale, the centre's director.

MEDICINE

Oncology burnout

Although 83% of US oncologists report career satisfaction, about 45% experience emotional exhaustion or other symptoms of burnout, says a study presented on 2 June at the meeting of the American Society of Clinical Oncology in Chicago, Illinois. The 2012–13 survey of about 1,500 oncologists found a link between burnout and high patient volume. Academic oncologists spend more time with patients and less on research than in the past, says lead author Tait Shanafelt, a haematologist and oncologist at the Mayo Clinic in Rochester, Minnesota. He suggests that early-career academic oncologists need to preserve their research time.

AWARDS

Prizes for the young

US researchers under the age of 42 will be able to vie for one of three annual unrestricted awards of US\$250,000 in life sciences, chemistry, and physical sciences and engineering, the Blavatnik Family Foundation in New York and the New York Academy of Sciences (NYAS) announced on 3 June. "We want to highlight young researchers who are doing such extraordinary and innovative work that it will incentivize other young researchers," says NYAS president Ellis Rubinstein. Nominations from US research universities and institutions, national labs and academic medical centres will be accepted from October to December 2013. NYAS council members may nominate industry researchers.

MORTAR FLOWERS

The art of remembering.

BY JESSICA MAY LIN

Sometimes in the morning, a single gull would cry, after the mortar shells had rained all night and spilt blood trickled down the alley walls into the sunbaked asphalt.

The Cement Florist boiled jars of coloured resin in the crumbling kitchen of his third-floor apartment, which overlooked the warships in the harbour. He bit a cigar between his teeth as he spooned hot resin out of its jar and let it fall back, occasionally glancing over his shoulder at the neatly made bed with its blue-and-pink-striped quilt.

He had awoken in the middle of the night to gunshots in the cul-de-sac. Another execution. It was at times like these, when he lay alone in the dark and the screams ate into his mind, that he missed her most.

Drawing his brown leather jacket around his shoulders, he set one of his jars under his arm and locked the door.

The alley was filled with cold, pale faces. Eyes open, staring lifelessly past him at the empty mustard gas canisters rolling in the shadows. The concrete had been blasted away by mortar bombs, leaving spiralling, blotched scars that decorated the pavement like bruises.

The Cement Florist opened a jar of hot yellow resin, honeyed vapours rising out of the glass. He slowly poured the contents into the whorled contours of a mortar scar.

Achillea millefolium. The bloodwort flower, once used on the battlefield to staunch a soldier's bleeding wounds.

The fires in the Juku Ghetto had finally died, taking the rotting tenements with them.

The Cement Florist stood under the overhang of a destroyed brothel, carrying his jars. The prostitutes glared at him with accusing eyes from where they huddled in the ruins, neglected, lace garters ripped and nails long.

Street urchins ducked in the rain, hugging to their chests the spokes of a broken chandelier they'd hauled out of the river after last night's flood.

➔ NATURE.COM

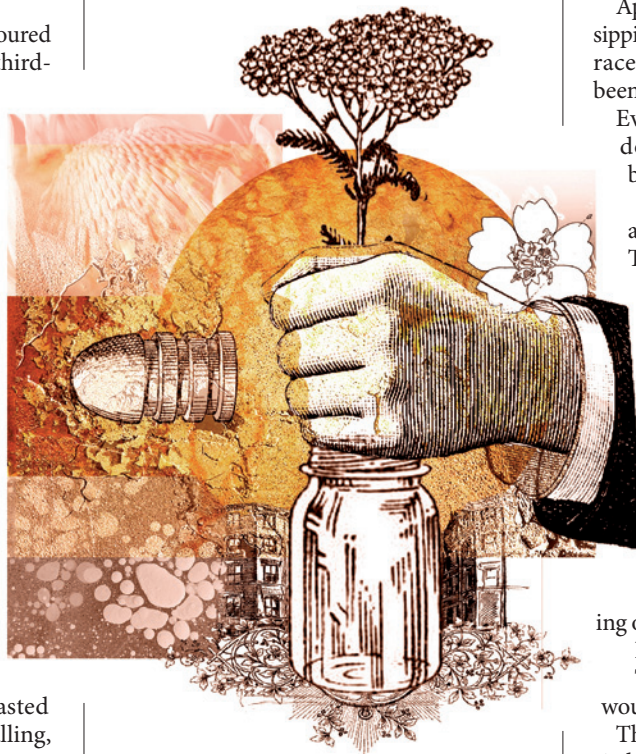
Follow Futures:

Twitter @NatureFutures

Facebook go.nature.com/mtoodm

This was where he had first met her, when he sold flowers out of his rusted truck

to the working men, for their sweethearts. Things had been different then. Lovers walked with their heads up, and children didn't fight each other with sticks. There hadn't been the pasty smell of ashes, which drifted down on the city like snowflakes.



He stepped into the charred street, his face streaked with rain and tears, and fell to his knees. He filled the concrete scars with blue resin, for the urchins' dirty scarves, wrapped around defunct mortar shells that they'd painted into dolls.

Myosotis scorpioides. The true forget-me-not, for children from whom the war had cruelly robbed their innocence, shivering in the cold and forgotten.

He used to walk with her in the Hanging Gardens, which now hung limp, brown and wilted from the mustard gas. Waffle crumbs still littered the marble walkways where young lovers once walked through the dappled sunlight, licking ice cream cones.

He'd read a story in the newspaper last week about a boy and a girl who tried to escape on the long bridge that led from the gardens out of the city. The snipers found them before they could taste freedom. Their bodies still lay entwined in the dust, where nobody had bothered to retrieve them.

Standing in the dry shadows of limp, dead ivy, the Cement Florist wondered about what could've been — if he had taken her hand and run. If they would be together right now, in this life or the next.

He sighed.

Apple pies and warm Saturday mornings sipping coffee in bed, watching sailing-boats race in the harbour. All these things had been stolen from his fingertips.

Eventually he set the jar of white resin down on the ground, and filled the blotched concrete.

Asphodelis aestivus. The summer asphodel, the flower of the underworld. They say that in the Silent Meadow — the place where all lovers are eventually united — it grows in soft fields, slowly bending in a nonexistent breeze. That's where he would meet her.

Back in his apartment, the Cement Florist sat down on the edge of the bed. He lifted the quilt with his blistered hands and breathed in her warm, lavender scent.

The hands of the brass clock hanging over the sink moved onto the hour.

He looked at the door.

They came when he had known they would.

The gloved men with cold faces, who carried rifles and ordered him to come outside into the street.

He followed them in silence, and thought of white flowers in a sweet-scented field, when they drew a knife across his neck and lay him down on the pavement to bleed.

His heartbeat was the last thing he heard, the sun warm against his skin, when he exhaled for the last time into the musty evening air.

His blood swirled into the concrete scar where a mortar shell had fallen that morning — a bright, flowing red.

Protea cynaroides. The king protea, the oldest flower in the world. One of great strength and courage — for a man who devoted himself to changing suffering into art, to making beauty where it no longer exists, even if no one will ever see it. ■

Jessica May Lin is a student at the University of California, Berkeley. Her work is also forthcoming in *Daily Science Fiction*. Her website is jessicamaylin.com.

JACEY

Properties of native brain α -synuclein

ARISING FROM T. Bartels, J. G. Choi & D. J. Selkoe *Nature* **477**, 107–110 (2011)

α -Synuclein is an abundant presynaptic protein that binds to negatively charged phospholipids^{1,2}, functions as a SNARE-complex chaperone³ and contributes to Parkinson's disease pathogenesis^{4,5}. Recombinant α -synuclein in solution is largely unfolded and devoid of tertiary structure^{6–11}, but Bartels *et al.*¹² have proposed that native α -synuclein purified from human erythrocytes forms a stably folded, soluble tetramer that resists aggregation. By contrast, we show here that native α -synuclein purified from mouse brain consists of a largely unstructured monomer, exhibits no stable tetramer formation, and is prone to aggregation. The native state of α -synuclein is important for understanding its pathological effects as a stably folded protein would be much less prone to aggregation than a conformationally labile protein. There is a Reply to this Brief Communication Arising by Bartels, T. & Selkoe, D. J. *Nature* **498**, <http://dx.doi.org/10.1038/nature12126> (2013).

We examined native α -synuclein from brain, the most relevant organ for understanding neurodegeneration. Separation of mouse brain homogenates into soluble and membrane fractions revealed that during ultracentrifugation, most α -synuclein partitioned into cytosol fractions similar to complexins, but different from membrane proteins such as cysteine string protein (CSP)- α and SNAP25 (Fig. 1a, b). Using gel filtration, we analysed the size of native α -synuclein in brain cytosol and of recombinant myc-epitope-tagged human α -synuclein, purified without boiling or detergents³. Both α -synucleins eluted in a single peak with an apparent molecular mass of ~ 63 kDa (Fig. 1c–f), close to that predicted for a folded tetramer¹².

These results seem to confirm that α -synuclein forms a stable tetramer in solution. However, dynamic or unstructured states of a protein may increase its hydrodynamic radius and apparent molecular mass

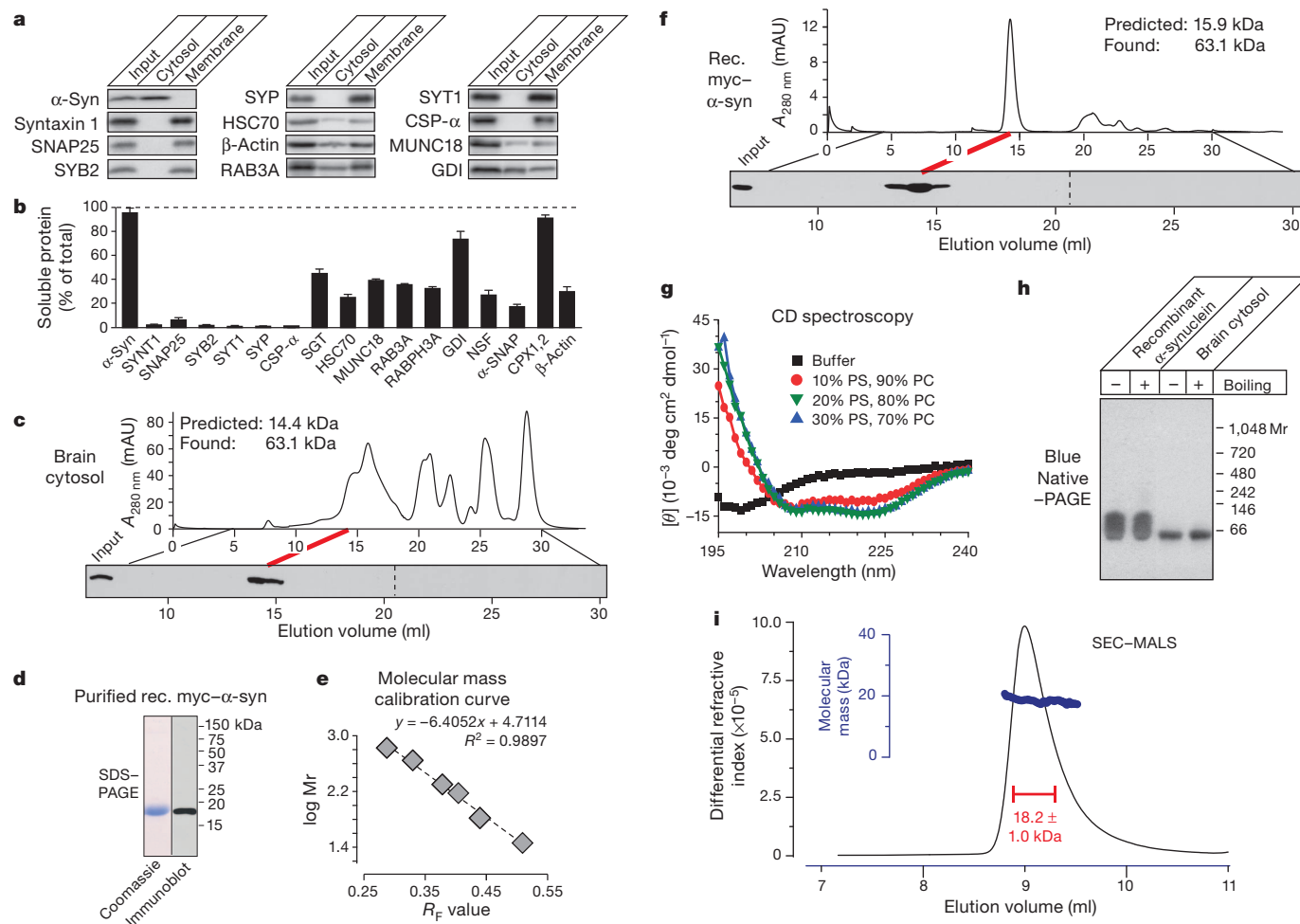


Figure 1 | Recombinant α -synuclein and brain α -synuclein in cytosol are monomeric. **a**, **b**, Immunoblotting analysis of mouse brain homogenate (input), cytosol and membranes (**a**), and quantification of protein levels (**b**; means \pm s.e.m.; $n = 3$). **c**, Native mouse brain α -synuclein (375 μ g) elutes as an apparent tetramer during gel filtration on a Superdex 200 column (top), as analysed by α -synuclein immunoblotting (bottom). mAU, milli absorbance unit. **d**, Analysis of purified recombinant myc-epitope-tagged α -synuclein (rec. myc- α -syn) by SDS-PAGE and immunoblotting. **e**, Molecular mass calibration curve for gel filtration (R_F = migration distance of proteins versus total running

distance; y axis = logarithm of molecular protein mass (Mr)). **f**, Recombinant myc-tagged human α -synuclein (16 μ g) also elutes as an apparent tetramer during gel filtration. **g**, Circular dichroism spectroscopy shows that recombinant α -synuclein (10 μ g) is unstructured in solution and becomes α -helical upon liposome binding. PC, phosphatidylcholine; PS, phosphatidylserine. Molar protein-to-lipid ratio, 1:530; θ = molar ellipticity. **h**, Recombinant (0.5 μ g) and α -synuclein in brain cytosol (12 μ g total protein) run as apparent tetramers on blue native gels without boiling or after boiling for 5 min. **i**, SEC-MALS reveals that recombinant α -synuclein (0.5 mg) is monomeric.

during gel filtration. Indeed, circular dichroism spectroscopy showed that recombinant α -synuclein lacked detectable secondary structure, but became α -helical upon membrane binding (Fig. 1g). Consistent with the gel-filtration analysis, both native and recombinant α -synuclein migrated as a single band of ~ 65 kDa on blue native gels. Notably, however, both recombinant and native α -synuclein still migrated at that apparent size after boiling, which disrupts secondary and tertiary structures, arguing against a folded multimer (Fig. 1h). Furthermore, size-exclusion chromatography coupled with multi-angle laser-light scattering (SEC-MALS) revealed that recombinant α -synuclein was monomeric (Fig. 1i). As native α -synuclein in brain cytosol and recombinant α -synuclein behave identically in gel filtration and blue native gel-electrophoresis experiments, the SEC-MALS demonstration that recombinant α -synuclein is monomeric suggests that native brain α -synuclein in cytosol is also monomeric.

We next tested whether native brain α -synuclein is still monomeric even when purified. We purified α -synuclein from mouse brain without detergents or denaturing conditions (purity $>90\%$; Fig. 2a). Mass spectrometry showed that native brain α -synuclein was substantially larger than predicted (measured mass, $16,408 \pm 894$ Da ($n = 3$); predicted mass, 14,485 Da). The increased mass is partly due to amino-terminal acetylation of brain α -synuclein^{12,13} (Fig. 2b). SEC-MALS revealed that freshly purified native α -synuclein was again predominantly monomeric (Fig. 2c). We also observed a plateau along the left shoulder of the main SEC-MALS peak with a mass of ~ 58 kDa that

contained little detectable α -synuclein ($<5\%$ by immunoblotting), and whose observed molecular mass is inconsistent with a putative tetramer. Circular dichroism spectroscopy showed a largely random-coil conformation (34–59%) with α -helical contributions (21–24%; Fig. 2d). Purified α -synuclein aggregated in a time-dependent manner, with a relative increase in overall secondary structure as observed by circular dichroism spectroscopy (Fig. 2d), and the appearance of larger particles as uncovered by dynamic light scattering (Fig. 2e).

Our data show that native brain α -synuclein primarily consists of an unstructured monomer, but readily aggregates in a time-dependent manner. This conclusion was demonstrated both for unpurified α -synuclein as a component of brain cytosol (Fig. 1), and for purified α -synuclein in solution (Fig. 2c). Purified brain α -synuclein — analysed here for the first time — carries significant post-translational modifications (Fig. 2b), which do not, however, seem to alter its folding, as the biophysical properties of recombinant unmodified α -synuclein and native modified α -synuclein were similar (Figs 1 and 2). The differences between our results with brain α -synuclein and those obtained with erythrocyte α -synuclein¹² may be due to erythrocyte-specific post-translational modifications, or to time-dependent multimerization/aggregation of erythrocyte α -synuclein that may have been overlooked. Indeed, the circular dichroism spectrum of erythrocyte α -synuclein¹² is similar to that of purified brain α -synuclein after 75 h incubation (Fig. 2d). Independent of which explanation will account for the differences in results obtained with brain and erythrocyte

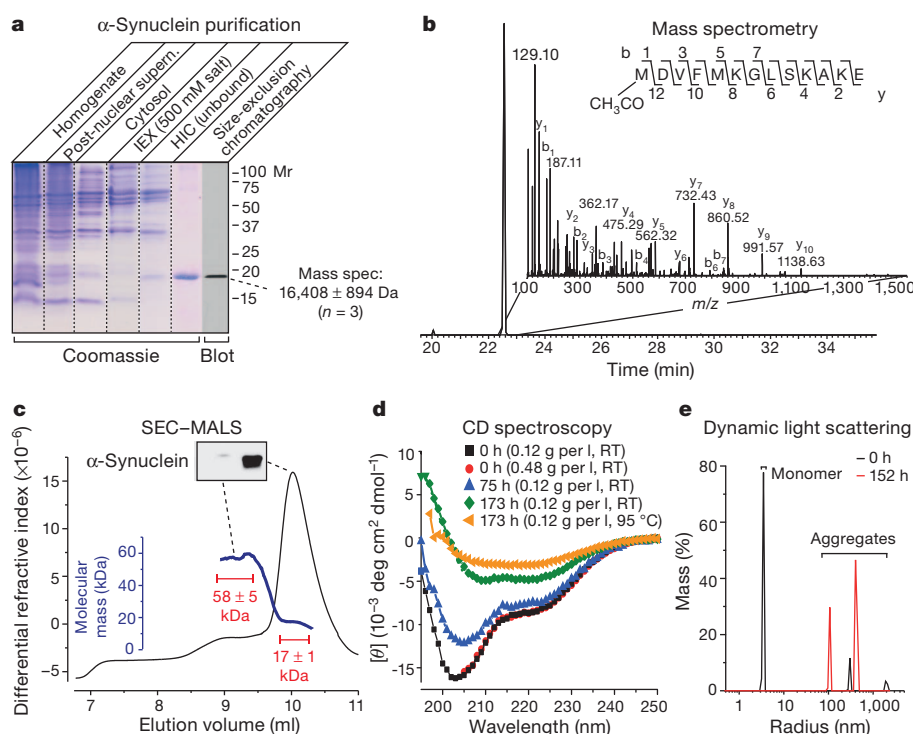


Figure 2 | Purified native brain α -synuclein is predominantly an unstructured monomer that aggregates in a time-dependent manner.

a, SDS-PAGE analysis of five stages of α -synuclein purification from mouse brain. IEX, anion exchange chromatography; HIC, hydrophobic interaction chromatography. Purified α -synuclein was also analysed by immunoblotting and mass spectrometry as shown. **b**, Mass spectrometry analysis reveals N-terminal acetylation of native α -synuclein. Shown is an extracted ion chromatogram of the N-terminally acetylated α -synuclein peptide. Inset, tandem MS spectrum containing the sequence of the N-terminal peptide and identified b and y ions. **c**, SEC-MALS shows that purified brain α -synuclein

(150 μ g) is largely monomeric (main peak with a mass of 17 ± 1 kDa), but includes a minor component (plateau along the left shoulder with a mass of 58 ± 5 kDa) that contains little detectable α -synuclein (see immunoblot in boxed region). Calculated masses were extracted from marked areas. **d**, Circular dichroism spectroscopy of freshly purified brain α -synuclein (0.12 g per l = 7.5 μ M) shows mainly disordered conformations that progressively acquire structured conformations as a result of time- and temperature-dependent aggregation. RT, room temperature. **e**, Purified brain α -synuclein (0.12 mg ml⁻¹) rapidly aggregates as measured by dynamic light scattering immediately (0 h) or 152 h after purification.

α -synuclein, the conformationally labile state of native brain α -synuclein documented here provides a potential explanation for why α -synuclein is susceptible to pathological aggregation as observed in multiple neurodegenerative disorders^{4,5}.

Methods

α -Synuclein was purified from mouse brain cytosol (obtained from brain homogenates by ultracentrifugation at 280,000g_{av}) by sequential chromatography on Q sepharose (elution at 0.3–0.5 M NaCl, 20 mM Tris-HCl, pH 7.4), phenyl sepharose (flow-through in 1 M (NH₄)₂SO₄) and Superdex-200 10/300GL SEC-MALS was performed on a WTC-030S5 column (Heleos OptiLab instruments, Wyatt Technology). Circular dichroism spectra were measured in 25% PBS on an Aviv CD Spectrometer and deconvolved (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) with Contin-4 and -7 reference sets. Mass spectrometry was performed on purified α -synuclein or α -synuclein-containing gel pieces digested with Glu-C and Protease Max (Promega, using standard procedures)¹⁴. All other methods have been described previously³.

Jacqueline Burré^{1*}, Sandro Vivona^{1,2,3,4*}, Jiajie Diao^{1,2,3,4},
Manu Sharma¹, Axel T. Brunger^{1,2,3,4,5} & Thomas C. Südhof^{1,4,5,6}

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA.

email: tcs1@stanford.edu

²Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

³Department of Photon Science, Stanford University School of Medicine, Stanford, California 94305, USA.

⁴Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, California 94305, USA.

⁵Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA.

⁶Department of Psychiatry, Stanford University School of Medicine, Stanford, California 94305, USA.

*These authors contributed equally to this work.

Received 10 January 2012; accepted 28 March 2013.

- Davidson, W. S., Jonas, A., Clayton, D. F. & George, J. M. Stabilization of α -Synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273**, 9443–9449 (1998).
- Jo, E., McLaurin, J., Yip, C. M., St George-Hyslop, P. & Fraser, P. E. α -Synuclein membrane interactions and lipid specificity. *J. Biol. Chem.* **275**, 34328–34334 (2000).
- Burré, J. *et al.* α -Synuclein promotes SNARE-complex assembly *in vivo* and *in vitro*. *Science* **329**, 1663–1667 (2010).
- Martin, I., Dawson, V. L. & Dawson, T. M. Recent advances in the genetics of Parkinson's disease. *Annu. Rev. Genomics Hum. Genet.* **12**, 301–325 (2011).
- Devine, M. J., Gwinn, K., Singleton, A. & Hardy, J. Parkinson's disease and α -synuclein expression. *Mov. Disord.* **26**, 2160–2168 (2011).
- Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A. & Lansbury, P. T. Jr. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**, 13709–13715 (1996).
- Kim, J. Evidence that the precursor protein of non-A beta component of Alzheimer's disease amyloid (NACP) has an extended structure primarily composed of random-coil. *Mol. Cells* **7**, 78–83 (1997).
- Uversky, V. N., Lee, H. J., Li, J., Fink, A. L. & Lee, S. J. Stabilization of partially folded conformation during α -synuclein oligomerization in both purified and cytosolic preparations. *J. Biol. Chem.* **276**, 43495–43498 (2001).
- Chandra, S., Chen, X., Rizo, J., Jahn, R. & Südhof, T. C. A broken α -helix in folded α -synuclein. *J. Biol. Chem.* **278**, 15313–15318 (2003).
- Lokappa, S. B. & Ulmer, T. S. α -Synuclein populates both elongated and broken helix states on small unilamellar vesicles. *J. Biol. Chem.* **286**, 21450–21457 (2011).
- Fauvet, B. *et al.* α -synuclein in the central nervous system and from erythrocytes, mammalian cells and *E. coli* exists predominantly as a disordered monomer. *J. Biol. Chem.* **287**, 15345–15364 (2012).
- Bartels, T., Choi, J. G. & Selkoe, D. J. α -Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **477**, 107–110 (2011).
- Maltese, A. S., Ying, J. & Bax, A. Impact of N-terminal acetylation of α -synuclein on its random coil and lipid binding properties. *Biochemistry* **51**, 5004–5013 (2012).
- Johnson, H. & Eyers, C. E. Analysis of post-translational modifications by LC-MS/MS. *Methods Mol. Biol.* **658**, 93–108 (2010).

Author Contributions J.B., S.V., J.D. and M.S. performed the experiments. All authors planned and analysed the experiments and wrote the paper.

Competing Financial Interests Declared none.

doi:10.1038/nature12125

Bartels & Selkoe reply

REPLYING TO J. Burré *et al.* *Nature* **498**, <http://dx.doi.org/10.1038/nature12125> (2013)

In disagreeing with our report that native α -synuclein occurs physiologically as an α -helically folded tetramer in neural and erythroid cells¹, Burré *et al.*² conclude instead that 'native brain α -synuclein' consists of a largely unstructured monomer. They make two implications about our paper that are inaccurate: (1) that our findings pertained only to erythrocyte α -synuclein (we reported multiple experiments on neural cells); and (2) that we concluded that cellular α -synuclein is a stable tetramer under all conditions (we did not use the term 'stable', and we observed monomers and some other oligomers in normal cells (e.g., Fig. 1d of ref. 1)). Indeed, we emphasized the need to discover "compounds that ... could kinetically stabilize native tetramers and prevent pathogenic α -synuclein aggregation". Although the data in our report suggest that tetramers are the predominant native species, tetramers and other oligomers arise from monomers, so there must be an equilibrium between monomeric and oligomeric forms in cells. Pathogenic events (e.g., mutations) could alter this equilibrium, and some therapeutic compounds could potentially re-establish it, as we explicitly suggested¹.

Most findings in Fig. 1 of Burré *et al.*² confirm previous reports (including ours¹) that recombinant α -synuclein is an unfolded monomer of ~14 kDa but migrates anomalously at ~60 kDa in gel filtration (their Fig. 1f), presumably owing to the large hydrodynamic radius of an extended monomer. We had stated that this made "gel filtration an

unreliable indicator [of mass] and therefore [it was] not used here"¹. That recombinant α -synuclein becomes α -helical upon binding phospholipid vesicles (their Fig. 1g) was also long known³ and observed by us¹. The key difference from our work regards their data on the folding and assembly state of native α -synuclein (their Fig. 2). We believe these data are less in disagreement with our conclusions than the authors suggest. First, they show by size-exclusion chromatography coupled with multi-angle laser-light scattering (SEC-MALS) the existence of small amounts of α -synuclein tetramer (58.5 kDa) in their natively purified brain preparation (their Fig. 2c). Then, their Fig. 2d shows circular dichroism spectra of purified brain α -synuclein that display a mixture of unfolded (34–59%) and α -helically folded (21–24%) protein, a clear structural difference from recombinant α -synuclein, which is all unfolded (their Fig. 1g, 'buffer'). Their findings are not entirely incompatible with our paper, as we had stated that helical tetramers were the predominant physiological species but variable amounts of monomers and other oligomers were observed¹.

Given that even the helically folded tetramer suggested by us¹ (and others^{4,5}) contains only about 50% helical structure (as the regions around amino acid 50 form structured loops and the carboxy terminus is conformationally mobile), the fact that their circular dichroism spectrum contains ~24% helical conformation suggests that up to half of their

brain α -synuclein sample is folded and the other half is unfolded. The latter result raises the possibility of either differences in tetramer:monomer equilibria between their (murine brain) and our (human erythrocyte or neuroblastoma) samples or a partial denaturation of the brain sample during purification. Interestingly, room-temperature incubation of their unfolded monomeric/partly folded tetrameric sample led to overall loss of circular dichroism spectral intensity (by $\sim 50\%$), probably due to aggregation and precipitation of some of the protein out of solution, and a relative increase in helical content of the protein remaining in solution (their Fig. 2d, green). The authors correctly indicate that now their spectrum of purified brain α -synuclein is similar to our spectrum of purified erythrocyte α -synuclein. They say this conversion indicates that “purified α -synuclein aggregated in a time-dependent manner, with a relative increase in secondary structure”, but using the term ‘aggregation’ for this helical change is different from the widely studied pathogenic aggregation of α -synuclein that involves a conversion to a β -sheet-rich structure. It was the latter type of aggregation that we showed native α -synuclein to be resistant to (Fig. 3d of ref. 1). Burré *et al.*² only observed loss of α -helical structure after heating brain α -synuclein to 95 °C (their Fig. 2d, orange), a condition that similarly led to denaturation of our purified α -synuclein helical tetramers (Supplementary Fig. 11 of ref. 1) and thus does not disprove our conclusion that native helical α -synuclein does not readily aggregate under physiological conditions.

The loss of overall circular dichroism signal accompanied by an increase in α -helical spectral components that Burré *et al.*² show in Fig. 2d could be interpreted in two ways: (1) some sample precipitation occurs, and at the same time the remaining soluble α -synuclein becomes increasingly α -helically folded (such an event could be interpreted as the refolding of a partially denatured protein); or (2) the monomeric, unfolded portion of the mixture (their Fig. 2c) aggregates and precipitates out of solution (their Fig. 2e), whereas the helically folded, apparently tetrameric component (their Fig. 2c) stays unaltered in solution and provides the circular dichroism signal. The latter interpretation would be consistent with our hypothesis that destabilization of helical tetramers into unfolded monomers in cells may precede pathological α -synuclein aggregation¹. In summary, the difference between their purified brain α -synuclein and our purified erythrocyte and neuroblastoma α -synuclein seems to be the relative abundance of the aggregation-resistant helical material at the time of initial analysis.

Even though the dynamic light scattering data of Burré *et al.*² in Fig. 2e imply an increasing amount of aggregates (in agreement with the partial precipitation suggested in their Fig. 2d), no conclusion

about the amount of remaining monomers/tetramers in the sample can be drawn from this, given the inability of dynamic light scattering to detect small particles if sufficient amounts of large particles are present in the mixture.

Collectively, the data of Burré *et al.*² show the existence of some helically folded, apparently tetrameric (58.5 kDa) protein in purified α -synuclein isolated from normal mouse brain, although in their hands, this constitutes only half (by their Fig. 2d) or a minor portion (by their Fig. 2c) of their total protein immediately after purification and only becomes the major species upon incubation over time (their Fig. 2d). Given that the two studies are therefore debating the relative proportion under native conditions of helically folded tetramers, not their existence *per se*, we believe it is reasonable to pursue attempts to stabilize helically folded native α -synuclein tetramers as an approach to reducing the pathological aggregation of monomers. In light of our findings in Bartels *et al.*⁴ and in an extensive α -synuclein crosslinking analysis in intact neurons and other cells⁶, the combined recent data support the hypothesis that physiological α -synuclein occurs in cells in an oligomeric (principally tetrameric) state in the cytosol^{1,4,5,6} which is in equilibrium with unfolded monomers.

This Reply is written by two out of three of the authors from the original paper¹. J. G. Choi left the laboratory for Graduate School in 2011.

Tim Bartels¹ & Dennis J. Selkoe¹

¹Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.

email: dselkoe@rics.bwh.harvard.edu

1. Bartels, T., Choi, J. G. & Selkoe, D. J. α -Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **477**, 107–110 (2011).
2. Burré, J. *et al.* α -Synuclein in brain cytosol is monomeric. *Nature* **498**, <http://dx.doi.org/10.1038/nature12125> (2013).
3. Davidson, W. S., Jonas, A., Clayton, C. F. & George, J. M. Stabilization of α -Synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273**, 9443–9449 (1998).
4. Wang, W. *et al.* A soluble α -synuclein construct forms a dynamic tetramer. *Proc. Natl Acad. Sci. USA* **108**, 17797–17802 (2011).
5. Westphal, C. H. & Chandra, S. S. Monomeric synucleins generate membrane curvature. *J. Biol. Chem.* **288**, 1829–1840 (2013).
6. Dettmer, U., Newman, A. J., Luth, E. S., Bartels, T. & Selkoe, D. *In vivo* crosslinking reveals principally oligomeric forms of α -synuclein and β -synuclein in neurons and non-neural cells. *J. Biol. Chem.* **288**, 6371–6385 (2013).

doi:10.1038/nature12126

Conserved regulatory elements in AMPK

ARISING FROM B. Xiao *et al.* *Nature* **472**, 230–233 (2011)

The AMP-activated protein kinase (AMPK), an $\alpha\beta\gamma$ heterotrimeric enzyme, has a central role in regulating cellular metabolism and energy homeostasis¹. The α -subunit of AMPK possesses the catalytic kinase domain, followed by a regulatory region comprising the auto-inhibitory domain (AID) and α -linker^{2,3}. Structural and biochemical studies suggested that AID is central to mammalian AMPK regulation⁴; however, this notion has been challenged recently by Xiao *et al.* on the basis of their active AMPK structure (Protein Data Bank accession 2Y94)⁵. On close inspection, however, we found that the α -subunit regulatory region was incorrectly built in their model, and our rebuilt model suggests a universal occurrence of the AID domain in AMPKs; we have also identified a novel regulatory motif that is essential for AMPK regulation.

The AID domain from *Schizosaccharomyces pombe* AMPK-like kinase folds into a non-canonical UBA conformation comprising three α -helices⁴. However, Xiao *et al.*⁵ reported in their structure of an active AMPK containing the kinase domain and flanking regulatory region that the AID region was found to be disordered. In addition,

their bioinformatic analysis suggested that the AID region in vertebrate AMPKs may adopt a four helical conformation. Thus, they proposed that the vertebrate AMPK does not contain an AID domain as in the yeast AMPK. We determined the crystal and NMR structures of two AID fragments from rat $\alpha 1$ and human $\alpha 2$ subunits, respectively, showing that each of the AID domains comprises only three α -helices stabilized by highly conserved hydrophobic residues (Fig. 1a, b). We found that the overall structures of the isolated mammalian AIDs share a markedly similar conformation to that of the yeast orthologue, indicating that AID is highly conserved throughout evolution from yeast to mammal.

AMPK is characterized by its ability to be regulated by binding of adenine nucleotides to the γ -subunit^{6,7}. One important feature of the structure of Xiao *et al.*⁵ is that a small α -subunit segment ($\alpha 1$ 373–382, termed the α -hook) was modelled to interact with the γ -subunit (Fig. 1a). However, His 376 of the α -subunit (α His 376) is unfavourably nested in a positive cave and the adjacent phosphorylated Thr 377 (pThr377) is positioned in a negative pocket. Thus, the potentially mis-represented AID structure together with the possibly incorrect structure of the α -hook prompted us to re-examine the crystallographic data of the assignment of Xiao *et al.*⁵. The α -hook was previously back-traced from the α -subunit carboxy-terminal domain; however, there is a discontinuity of electron density between residues α Asn 382 and α Ala 394. We thus rebuilt the model, in that the strong side-chain feature is reassigned as α Arg 363 (previously assigned to pThr 377) and α Glu 362 takes the place of α His 376 (Fig. 1c, d). The remaining residues of the α -regulatory region were manually built, and the amino acid sequence of the AID helix $\alpha 3$ and the following α -linker is off-registered by more than 10 amino acids relative to the previous structure (Fig. 1a). Notably, the two helical regions following the kinase domain are respectively reassigned, corresponding to AID helices $\alpha 1$ and $\alpha 3$, and they adopt similar conformation in isolated AID and trimeric AMPK structures (Appendix Fig. 1). The orientation of AID in the rebuilt model is consistent with our conformational switch model for

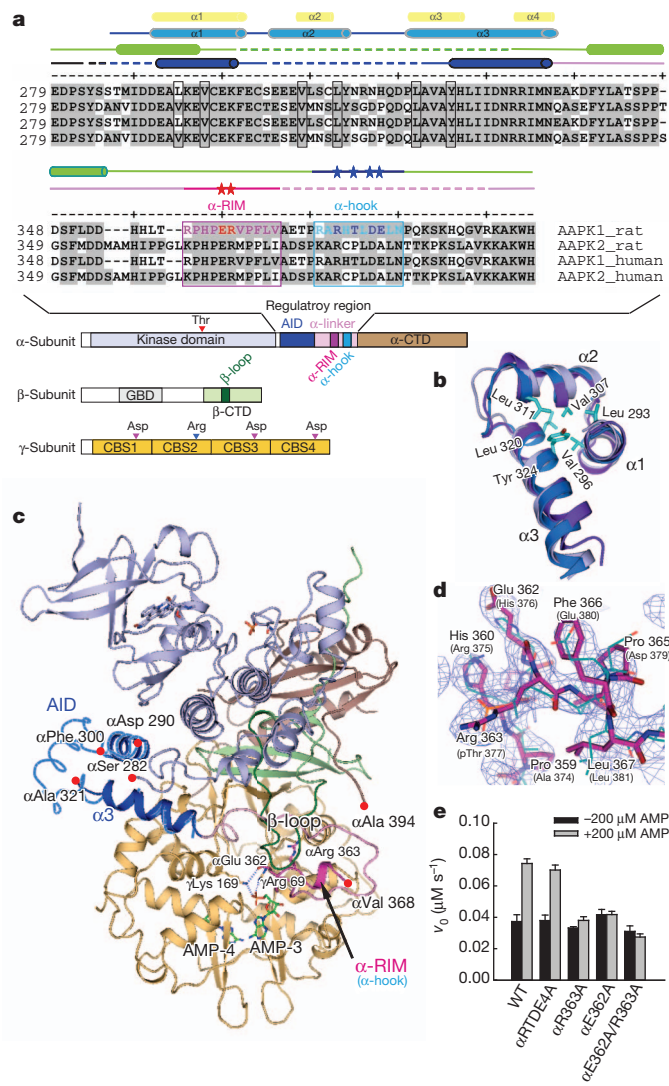


Figure 1 | Conserved AID and α -RIM for AMPK allosteric activation.

a, Sequence alignment of the α -subunit regulatory region from mammalian AMPKs. The newly identified α -RIM is boxed in magenta, and previous α -hook in cyan. Secondary structural elements in the predicted (yellow) and isolated (marine blue) AID structures and those in the active AMPK determined previously (green) and here (blue) are shown above the alignment. Key interacting residues in our α -RIM and previous α -hook are indicated by red and blue asterisks, respectively. Shown at the bottom is the schematic diagram of the $\alpha\beta\gamma$ heterotrimeric AMPK: kinase domain, light blue; AID, blue; α -linker, pink; α -RIM, magenta; α -hook, cyan; α -CTD, brown; β -CTD, green; β -loop, dark green; γ -subunit, light orange. CTD, C-terminal domain. **b**, Crystal structure of rat $\alpha 1$ -AID (marine blue) and solution structure of human $\alpha 2$ -AID (purple). The two isolated AID structures are superposed to the AID domain (light blue) in the *S. pombe* kinase domain-AID structure. The conserved hydrophobic core residues of rat $\alpha 1$ -AID are highlighted as cyan sticks. **c**, Graphical representations of the rebuilt active AMPK. The colour scheme follows the schematic diagram in **a**, with the two AMP molecules at sites 3 and 4 highlighted as green sticks. The rat $\alpha 1$ -AID (marine blue ribbon) is superposed to the putative AID helix $\alpha 1$ and $\alpha 3$. **d**, SA-omit map (contoured at 2.0σ) for the α -RIM (α -hook). Residues of the rebuilt and previous models are shown in magenta sticks and bluish green lines, respectively. **e**, The α -RIM has significant roles in allosteric activation of AMPK. The kinase assay was performed in the presence of 6.6 nM phosphorylated AMPK holoenzyme (wild type or mutants), 200 μ M SAMS, 1 mM ATP and 10 mM MgCl_2 , with or without addition of 200 μ M AMP (mean and s.e.m., $n = 3$).

AMPK regulation^{4,8}: AMP binding to the γ -subunit might transmit, via the α -linker, onto the AID, release AID from the kinase domain and ultimately activate AMPK. Thus, we consider that the AID domain is largely folded in the active AMPK, whereas the previously identified α -hook is disordered in the rebuilt model (Fig. 1c).

Instead of the α -hook, we identified a novel α -subunit motif (α 1 358–368) termed regulatory-subunit-interacting motif, α -RIM (Fig. 1c). The α Arg 363 penetrates into the pocket formed by a newly observed β -loop (human β 2 219–235) that was disordered in any other core structure^{8–12}. The γ -subunit contains four potential nucleotide-binding sites, of which sites 3 and 4 are important for AMPK allosteric stimulation by AMP^{8,13}. The acidic α Glu 362 interacts with the key γ -subunit residues Arg 69 and Lys 169 that are also in contact with the AMP molecule bound at site 3. Notably, these two important interacting residues in the α -RIM are highly conserved in mammals. To assess their significance, we generated point mutations on rat α 1 β 1 γ 1 trimer and examined their effects on AMPK allosteric activation^{4,14}. The wild-type holoenzyme, as well as the reported α -hook mutant α RTDE4A⁵, was activated about twofold by AMP (Fig. 1e). Mutating the two charged α -RIM residues α Glu 362 and α Arg 363 individually or simultaneously to Ala (α E362A, α R363A and α E362A/R363A) largely abolished the AMP-dependent activation. Similarly, the β -loop deletion mutant ($\Delta\beta$ -loop(218–228) on rat β 1) no longer responded to AMP concentration change. These data clearly demonstrate that the α -RIM in the proximity of site 3 has essential roles in AMPK allosteric regulation.

In summary, we have provided evidence to reaffirm our earlier report⁴ that AID is universally present in AMPK, and hereby rectify the misinterpretation of the AID structure and the α -hook feature by Xiao *et al.*⁵. Moreover, the essential role of the α -RIM (instead of the previously defined, disordered α -hook) for AMPK allosteric regulation is strongly supported by its interaction with the crucial γ -subunit site 3 and mutagenesis data.

Methods

The crystal of rat α 1-AID (284–336) was obtained with a reservoir solution containing 0.1 M HEPES, pH 7.5, 0.3 M magnesium sulphate, 36% isopropanol (v/v) at 4 °C. The NMR structure of human α 2-AID (282–339) was determined with ¹⁵N and ¹⁵N/¹³C proteins. The active AMPK structure was rebuilt with the diffraction data (2Y94.cif) downloaded from Protein Data Bank. The structural statistics are summarized in Appendix Tables 1–3, respectively. The site-specific mutations were generated by overlap PCR procedure and verified by DNA sequencing. AMPK was phosphorylated by CaMKK β and the activity was determined with SAMS peptide using a coupled assay.

Lei Chen^{1†}, Feng-Jiao Xin¹, Jue Wang¹, Jicheng Hu², Yuan-Yuan Zhang¹, Shuo Wan², Lu-Sha Cao¹, Chang Lu¹, Peng Li¹, S. Frank Yan³, Dietbert Neumann⁴, Uwe Schlattner^{5,6}, Bin Xia², Zhi-Xin Wang¹ & Jia-Wei Wu¹

¹MOE Key Laboratory of Protein Sciences and Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing

100084, China.

email: jiaweiwu@mail.tsinghua.edu.cn

²Beijing Nuclear Magnetic Resonance Center, College of Life Sciences, and College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.

³Molecular Design and Biostructure, Roche Pharma Research and Early Development China, Shanghai 201203, China.

⁴Cardiovascular Research Institute Maastricht, Maastricht University, 6200 MD Maastricht, The Netherlands.

⁵INSERM U1055, 38041 Grenoble Cedex 9, France.

⁶Laboratory of Fundamental and Applied Bioenergetics, University Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France.

[†]Present address: Vollum Institute, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA.

Received 2 July 2012; accepted 22 March 2013.

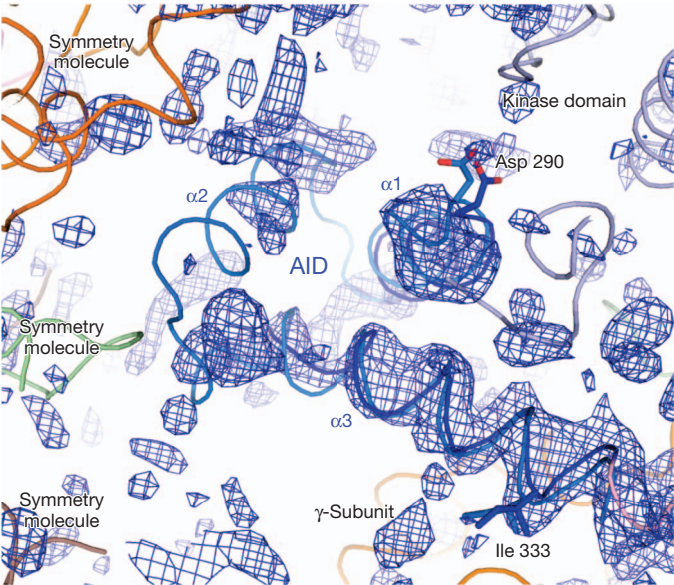
- Hardie, D. G., Ross, F. A. & Hawley, S. A. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nature Rev. Mol. Cell Biol.* **13**, 251–262 (2012).
- Crute, B. E., Seefeld, K., Gamble, J., Kemp, B. E. & Witters, L. A. Functional domains of the α 1 catalytic subunit of the AMP-activated protein kinase. *J. Biol. Chem.* **273**, 35347–35354 (1998).
- Pang, T. *et al.* Conserved α -helix acts as autoinhibitory sequence in AMP-activated protein kinase α subunits. *J. Biol. Chem.* **282**, 495–506 (2007).
- Chen, L. *et al.* Structural insight into the autoinhibition mechanism of AMP-activated protein kinase. *Nature* **459**, 1146–1149 (2009).
- Xiao, B. *et al.* Structure of mammalian AMPK and its regulation by ADP. *Nature* **472**, 230–233 (2011).
- Hardie, D. G., Carling, D. & Gamblin, S. J. AMP-activated protein kinase: also regulated by ADP? *Trends Biochem. Sci.* **36**, 470–477 (2011).
- Oakhill, J. S., Scott, J. W. & Kemp, B. E. AMPK functions as an adenylate charge-regulated protein kinase. *Trends Endocrinol. Metab.* **23**, 125–132 (2012).
- Chen, L. *et al.* AMP-activated protein kinase undergoes nucleotide-dependent conformational changes. *Nature Struct. Mol. Biol.* **19**, 716–718 (2012).
- Townley, R. & Shapiro, L. Crystal structures of the adenylate sensor from fission yeast AMP-activated protein kinase. *Science* **315**, 1726–1729 (2007).
- Amodeo, G. A., Rudolph, M. J. & Tong, L. Crystal structure of the heterotrimeric core of *Saccharomyces cerevisiae* AMPK homologue SNF1. *Nature* **449**, 492–495 (2007).
- Xiao, B. *et al.* Structural basis for AMP binding to mammalian AMP-activated protein kinase. *Nature* **449**, 496–500 (2007).
- Jin, X., Townley, R. & Shapiro, L. Structural insight into AMPK regulation: ADP comes into play. *Structure* **15**, 1285–1295 (2007).
- Oakhill, J. S. *et al.* β -Subunit myristoylation is the gatekeeper for initiating metabolic stress sensing by AMP-activated protein kinase (AMPK). *Proc. Natl Acad. Sci. USA* **107**, 19237–19241 (2010).
- Neumann, D., Woods, A., Carling, D., Wallimann, T. & Schlattner, U. Mammalian AMP-activated protein kinase: functional, heterotrimeric complexes by co-expression of subunits in *Escherichia coli*. *Protein Expr. Purif.* **30**, 230–237 (2003).

Author Contributions L.C., F.-J.X., J.W., Y.-Y.Z., L.-S.C. and C.L. performed the X-ray structural and mutational studies; J.H. and S.W. carried out NMR studies; Z.-X.W., U.S., D.N., B.X., S.F.Y. and P.L. contributed to discussions and manuscript writing; J.-W.W. led the team and wrote the paper. Coordinates have been deposited into the Protein Data Bank with accession codes 4F2L for the rat α 1-AID crystal structure and 2LTU for the human α 2-AID NMR solution structure, respectively.

Competing Financial Interests Declared none.

doi:10.1038/nature12189

Appendix



Appendix Figure 1 | SA-omit map (contoured at 2.0σ) for the AID region. The colour scheme of the rebuilt model follows that in Fig. 1c, and the isolated rat α1-AID is shown in marine blue. Two residues from AID helices α1 and α3, Asp 290 and Ile 333, are shown as stick representation. Upon superposition, helix α2 in the isolated AID is located between the γ-subunit CBS1 and β-CTD from a symmetry molecule. The loop between AID helices α2 and α3 slightly clashes with a loop from the symmetry β-CTD, but such loop regions can be easily accommodated with very minor adjustments in their conformations. Notably, there are discontinuous densities for the AID helix α2 in this omit map. Therefore, we think that the AID helix α2 in the active AMPK may not be fully disordered, but rather adopt slightly different conformations, which results in the poor electron density.

Appendix Table 1 | Data collection and refinement statistics for rat α1-AID crystal structure

Rat α1-AID (residues 284–336)	
Data collection*	
Space group	P222 ₁
Cell dimensions	
a, b, c (Å)	22.1, 36.3, 98.8
α, β, γ (°)	90, 90, 90
Resolution (Å)	50.0–1.5 (1.53–1.50)†
R _{merge} (%)	4.6 (26.3)
I/σ	36.4 (3.6)
Completeness (%)	99.5 (98.5)
Redundancy	3.6 (3.1)
Refinement	
Resolution (Å)	24.4–1.5
No. reflections	24,447
R _{work} /R _{free}	21.2/18.2
No. atoms	
Protein	524
Water	79
B-factors (Average)	
Protein	22.52
Water	20.43
r.m.s.d.	
Bond lengths (Å)	0.005
Bond angles (°)	0.902

r.m.s.d., root mean square deviation.
* The data set was collected from single crystal.
† Highest resolution shell is shown in parenthesis.

Appendix Table 2 | Experimental and structural statistics for the ensembles of 20 NMR structures of human α2-AID

Human α2-AID (residues 282–338)	
Distance constraints	
Intra-residue ($ i-j = 0$)	587
Sequential ($ i-j = 1$)	314
Medium ($2 \leq i-j \leq 4$)	277
Long-range ($ i-j \geq 5$)	190
Ambiguous	379
Total	1,747
Dihedral angle constraints	
φ	42
ψ	42
Total	84
Hydrogen bond constraints	
Structure statistics (20 structures)	
Violation statistics	
NOE violation (>0.3 Å)	0
Maximum NOE violation (Å)	0.16
Torsion angle violation ($>5^\circ$)	0
Energy	
Mean AMBER energy (kcal mol ⁻¹)	-2,230.8
Mean bond energy	33.7
Mean angle	122.5
Mean dihedral	550.7
Mean VDW	-423.3
Ramachandran plot analysis	
Most favoured regions (%)	90.1
Additional allowed regions (%)	8.7
Generously allowed regions (%)	0.8
Disallowed regions (%)	0.4
r.m.s.d. from mean structure*†	
Backbone atoms (Å)	0.52 ± 0.11‡
All heavy atoms (Å)	1.06 ± 0.10‡
Regular secondary structures (Å)*†	
Backbone atoms (Å)	0.31 ± 0.07‡
All heavy atoms (Å)	1.05 ± 0.12‡

NOE, nuclear overhauser effect; VDW, van der Waals.
* The average r.m.s.d. between the 20 structures of the lowest AMBER energies and the mean coordinates (± standard deviation).
† Calculated with PROCHECK_NMR.
‡ Residues 12–56 in AID were used in the calculation.

Appendix Table 3 | Refinement statistics for the active AMPK structure using diffraction data (2Y94.cif) downloaded from PDB

Heterotrimeric AMPK (rat α1, human β2, rat γ1)	
Refinement	
Resolution (Å)	29.53–3.24
No. reflections	19,619
R _{work} /R _{free}	24.70/28.40
No. atoms	6,605
Protein	6,524
Ligand/ion	81
B-factors (Average)	133.29
r.m.s.d.	
Bond lengths (Å)	0.014
Bond angles (°)	0.652